

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 2, 2003, 06:03:31 ; Search time 63 Seconds

(without alignments)

1015.243 Million cell updates/sec

Title: US-09-830-980-1

Perfect score: 2574

Sequence: 1 MQETDTEQATPHTIQARLV.....PDGSRVASGGKDKVIKLWAY 480

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A.Geneseq.101002:\*

- 1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*
- 2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*
- 3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*
- 4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*
- 5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*
- 6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*
- 7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*
- 8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*
- 9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*
- 10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*
- 11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*
- 12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*
- 13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*
- 14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*
- 15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*
- 16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*
- 17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*
- 18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*
- 19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*
- 20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*
- 21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*
- 22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*
- 23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Query	Score	Match	Length	ID	Description
1	2558.5	99.4	479	21	AA79678	Drosophila Notchle
2	2554.5	99.2	481	22	AB659486	Drosophila melanog
3	1521	59.1	485	22	AA68284	Amino acid sequenc
4	1520	59.1	485	22	AA68284	Human protein sequ
5	1520	59.1	485	22	AA68284	Human GTP-binding
6	1520	59.1	485	23	AB97306	Novel human protei
7	1452	56.4	484	22	AB68282	Amino acid sequenc
8	1335.5	51.9	471	21	AA623141	Arabidopsis thalia
9	1335.5	51.9	471	21	AA64521	Arabidopsis thalia
10	1335.5	51.9	473	21	AA623140	Arabidopsis thalia

11	1335.5	51.9	473	21	AA64520	Arabidopsis thalia
12	1335.5	51.9	490	21	AA623139	Arabidopsis thalia
13	1335.5	51.9	490	21	AA64519	Arabidopsis thalia
14	1058.5	41.1	514	16	AA685881	WD-40 domain-contg
15	819	31.8	261	22	AA68283	Amino acid sequenc
16	419	16.3	317	21	AA63186	Human secreted pro
17	419	16.3	334	22	AA68529	Human GTP-binding
18	419	16.3	334	23	AB97345	Novel human protei
19	415	16.1	361	22	AB68576	Drosophila melanog
20	412.5	16.0	700	22	AB60376	Drosophila melanog
21	399	15.5	330	22	AA63659	Human protein sequ
22	398.5	15.5	341	21	AA63874	Arabidopsis thalia
23	389	15.1	411	22	AB62260	Drosophila melanog
24	372.5	14.5	514	22	AA65223	Human protein sequ
25	372	14.5	323	21	AA63185	Gene 3 human secre
26	371	14.4	542	23	AB41760	Human ovarian anti
27	355	13.8	333	21	AA65554	Arabidopsis thalia
28	354	13.8	409	16	AA670002	OPDE 45 kba subuni
29	354	13.8	410	16	AA670005	OPDE 45 kba subuni
30	348	13.5	478	22	AA63784	Human polypeptide,
31	348	13.5	521	22	AB10141	Human cDNA SEQ ID
32	344	13.4	159	20	AA10919	Amino acid sequenc
33	343	13.3	407	22	AA63675	Human polypeptide,
34	340.5	13.2	409	16	AA68568	WD-40 domain-contg
35	335	13.0	358	22	AB65223	Drosophila melanog
36	331	12.9	584	22	ABG21351	Novel human diagno
37	330	12.8	423	22	AA64261	Human protein sequ
38	330	12.8	423	22	AA64262	Human protein sequ
39	327	12.7	388	21	AA64313	Human ORF3077
40	321	12.5	423	22	AA63290	Human protein sequ
41	320	12.4	423	20	AA62249	Human secreted pro
42	320	12.4	423	22	AA63068	Human secreted pro
43	320	12.4	423	22	AA63068	Human polypeptide
44	320	12.4	423	23	AB65777	Human polypeptide
45	320	12.4	447	22	AA64155	Human polypeptide

#### ALIGNMENTS

RESULT 1  
AA79678  
ID AA79678 standard; Protein; 479 AA.

AC AA79678;

DT 29-AUG-2000 (first entry)

DE Drosophila Notchless protein.

XX Notchless; Nle gene; Notch; signalling; neurodegenerative disease;  
cancer; diagnosis; cytosolic; neuroprotective; therapy.

OS Drosophila melanogaster.

Key Location/Qualifiers  
Domain 27...105

FT Misc-difference 67 /note= "Nle domain"

FT Misc-difference 67 /note= "sequence deduced from nucleotide sequence  
has an additional Lys residue between  
Lys-67 and Ser-68"

FT Misc-difference 282 /note= "encoded by CAA"

XX WC200026364-AL.

XX 11-MAY-2000.

XX 03-NOV-1999; 99WO-IB01891.

XX 03-NOV-1998; 98GB-0024045.

(EURO-) EURO MOLECULAR BIOLOGY LAB.

Cohen S, Bouwmeester A, Royet J;

WPI: 2000-365613/31.

N-PSDB; AAZ27739.

Novel Notchless protein and nucleic acids encoding them useful for treating and preventing cancer and neurodegenerative diseases

Claim 1; Page 44; 52pp; English.

The present sequence is that of Notchless, a novel protein of Drosophila. Notchless was identified in a screen for dominant modifiers of a Notch mutant phenotype in the Drosophila wing. The mutant dominantly suppressed the wing notching phenotype of Notchoid mutations, and the Notchless protein was shown to bind to the cytoplasmic domain of Notch. Notchless modified Notch signalling activity in a variety of Notch-dependent signalling processes in both Drosophila and Xenopus embryos. The Notchless protein has a novel highly conserved N-terminal domain followed by 9 WD40 repeats. Notchless, and nucleic acids encoding it, can be used in methods for the diagnosis and therapy of certain diseases, particularly cancer and neurodegenerative diseases (claimed). A Notchless mutant in a sensitised Notch genetic background is used in a claimed method for identifying compounds capable of modifying the levels of expression or activity of a Notch protein.

Sequence 479 AA;

Query Match 99.4%; Score 2558.5; DB 21; Length 479;

Best Local Similarity 99.8%; Pred. No. 1.1e-232;

Matches 479; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 MQETDEQATPHTIQARLVYTGEEAGPPIDLPAGITTOQLGLICNALLKNEATPYLFF 60

Db 1 MQETDEQATPHTIQARLVYTGEEAGPPIDLPAGITTOQLGLICNALLKNEATPYLFF 60

QY 61 VGEDEIKKSLDLDLASVDTEENVIVYQOAVFKVRPVTRCTSSMPGHAENVSLNF 120

Db 61 VGEDEIKKSLDLDLASVDTEENVIVYQOAVFKVRPVTRCTSSMPGHAENVSLNF 119

QY 121 PDGAHLASGSDTIVRLWDLNTPHFTCTGHKQWLVCSWAPDGKRLASGCKAGSI 180

Db 120 PDGAHLASGSDTIVRLWDLNTPHFTCTGHKQWLVCSWAPDGKRLASGCKAGSI 179

QY 181 DPETGQOKGRPLSGHKHINCLAWEPYHRDPCRLASASGDGDCRWDVKGQCLM 240

Db 180 DPETGQOKGRPLSGHKHINCLAWEPYHRDPCRLASASGDGDCRWDVKGQCLM 239

QY 241 GHTNAVTVRWGGAGLIYTSKDRVTVMWRAADGILCTFSGHAHWNVNIALSTDY 300

Db 240 GHTNAVTVRWGGAGLIYTSKDRVTVMWRAADGILCTFSGHAHWNVNIALSTDY 299

QY 301 GPFHPVKDRSKSHLSLSTEELQESALKRYQAVCPDEVESLVSCSDNTLYLRNNQ 360

Db 300 GPFHPVKDRSKSHLSLSTEELQESALKRYQAVCPDEVESLVSCSDNTLYLRNNQ 359

QY 361 ERMTHQNVNDYKSPDKLIASAFDLSVLRASDGOYMATFGRHVOAVTVWASAD 420

Db 360 ERMTHQNVNDYKSPDKLIASAFDLSVLRASDGOYMATFGRHVOAVTVWASAD 419

QY 421 SRLIVSGKSDTLKVMVSVQTKLAQELPGHADEVFGVMDPAGSRVSGGDKVKIK 480

Db 420 SRLIVSGKSDTLKVMVSVQTKLAQELPGHADEVFGVMDPAGSRVSGGDKVKIK 479

RESULT 2

ABB59486

ID ABB59486 standard; Protein; 481 AA.

XX

AC ABB59486;

XX

DT 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 5250.

XX Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI: 2001-656860/75.

XX N-PSDB; ABL03589.

XX New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions

XX Disclosure; SEQ ID NO 5250; 21pp + Sequence Listing; English.

The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-ABB72072).

The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

Sequence 481 AA;

Query Match 99.2%; Score 2554.5; DB 22; Length 481;

Best Local Similarity 99.6%; Pred. No. 2.7e-232;

Matches 479; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 MQETDEQATPHTIQARLVY-TGEEAGPPIDLPAGITTOQLGLICNALLKNEATPYLFF 59

Db 1 MQETDEQATPHTIQARLVSDTGEAGPPIDLPAGITTOQLGLICNALLKNEATPYLFF 60

QY 60 FVGEDEIKKSLDLDLASVDTEENVIVYQOAVFKVRPVTRCTSSMPGHAENVSLNF 119

Db 61 FVGEDEIKKSLDLDLASVDTEENVIVYQOAVFKVRPVTRCTSSMPGHAENVSLNF 120

QY 120 SPDGAHLASGSDTIVRLWDLNTPHFTCTGHKQWLVCSWAPDGKRLASGCKAGSI 179

Db 121 SPDGAHLASGSDTIVRLWDLNTPHFTCTGHKQWLVCSWAPDGKRLASGCKAGSI 180

QY 180 WDPETGQOKGRPLSGHKHINCLAWEPYHRDPCRLASASGDGDCRWDVKGQCLM 239

Db 181 WDPETGQOKGRPLSGHKHINCLAWEPYHRDPCRLASASGDGDCRWDVKGQCLM 240

QY 240 AGHTNAVTVRWGGAGLIYTSKDRVTVMWRAADGILCTFSGHAHWNVNIALSTDY 299

Db 241 AGHTNAVTVRWGGAGLIYTSKDRVTVMWRAADGILCTFSGHAHWNVNIALSTDY 300

QY 300 TGPFPHPVKDRSKSHLSLSTEELQESALKRYQAVCPDEVESLVSCSDNTLYLRNNQ 359

Db 301 TGPFPHPVKDRSKSHLSLSTEELQESALKRYQAVCPDEVESLVSCSDNTLYLRNNQ 360

QY 360 VERTGHQNVNDVYKSPDKSVRLMRASDFKSVRLMRASDQGYMATFRGHVQAVYTVAMSA 419  
 Db 361 VERTGHQNVNDVYKSPDKSVRLMRASDFKSVRLMRASDQGYMATFRGHVQAVYTVAMSA 420  
 QY 420 DSRLLVSGSKSTLKVNSVQTKKLAQELPCHADEVFGVDWAPDGSRVASGGKDKVIKWA 479  
 Db 421 DSRLLVSGSKSTLKVNSVQTKKLAQELPCHADEVFGVDWAPDGSRVASGGKDKVIKWA 480  
 QY 480 Y 480  
 Db 481 Y 481

## RESULT 3

AA068284 standard; Protein; 485 AA.

XX AA068284;

XX 09-JUL-2001 (first entry)

XX Amino acid sequence of a human Tzap gene cDNA clone.

XX Tzap; T cell activation; immune response; transplant rejection;  
 KW bone marrow transplantation; rheumatoid arthritis; lupus erythematosus;  
 KW multiple sclerosis; encephalitis; vasculitis; diabetes mellitus;  
 KW pancreatitis; gastritis; thyroiditis; malignant disorder; asthma;  
 KW lepromatosis; gastritis; skin tumour; adrenal tumour; lung tumour;  
 KW wound healing; growth disorder; inflammatory disease; infectious disease.

XX Homo sapiens.

XX Key Location/Qualifiers

XX Misc-difference 379 /note= "unspecified residue encoded by NTC"

XX WC0200132614-A2.

XX 10-MAY-2001.

XX 30-OCT-2000; 2000WO-EP10670.

XX 01-NOV-1999; 99US-0162675.

XX 25-FEB-2000; 2000US-0185016.

XX (UTKU/) UTKU N.

XX Utku N;

XX N-PSDB; 2001-316410/33.

XX Novel polynucleotide encoding Tzap protein involved in T cell

XX activation, useful for diagnosing and treating diseases involving T

XX cell activation, for treating organ transplantation rejection,

XX rheumatoid arthritis

XX Example; Page 56-57; 68pp; English.

XX The present sequence represents a Tzap protein. The Tzap gene is  
 CC involved in T cell activation, and in the modulation of immune  
 CC responses. Tzap polynucleotides and polypeptides are useful for  
 CC diagnosing or treating acute and chronic diseases involving T cell  
 CC activation and Th1 and Th2 immune response, for the treatment of acute  
 CC and chronic rejection of allo- and xeno organ transplants and bone  
 CC marrow transplantation, for the treatment of rheumatoid arthritis, lupus  
 CC erythematosus, multiple sclerosis, encephalitis, vasculitis, diabetes  
 CC mellitus, pancreatitis, gastritis, thyroiditis, for the treatment of  
 CC malignant disorders of T, B or NK cells, for the treatment of asthma,  
 CC lepromatosis, Helicobacter pylori associated gastritis or for the  
 CC treatment of skin tumours, adrenal tumours or lung tumours, wound  
 CC healing, growth disorders inflammatory and/or infectious diseases.

SQ Sequence 485 AA;  
 Query Match 59.1%; Score 1521; DB 22; Length 485;  
 Best Local Similarity 58.6%; Pred. No. 1.2e-134;  
 Matches 279; Conservative 73; Mismatches 118; Indels 6; Gaps 3;  
 QY 9 EATPHITQARLYVTTEEA---GPPIDLPAGITTTQOLGLICNALLKNEEATPYLFVFGED 64  
 Db 8 EAVARDVQRLVQFQDEGGQLLGSPFVDITPDRLQVNCNALLAQEDPLPLAEFVHDA 67  
 QY 65 EIKKSLETDLDASVDTENWIDIVYQPAVFKVPVTRCTSSMPGHAEAVSVLNFSPDGA 124  
 Db 68 EIVSSLGKLTLESQAVETEKVLDDIYQPAIFRVAVTRCTSSLEGHSEAVISVAFSPGK 127  
 QY 125 HLAGSGDITVRDLNLTETPHFTCTGKHQWVLCYSWAPDCKRLASCKAGSIIWDPET 184  
 Db 128 YLAGSGDITVRFNDLSTETPHFTCKGRHWVLSISWSPDKKLLASCKNGOILLWDPST 187  
 QY 185 GQKGRPLSGHKHINCLANEPYHRDPECKKLASAGSDGDCRIWDVKLGOCIMNTAGHTN 244  
 Db 188 GKQVGRTLAGHSKWITGLSWELPHANPECRYVASSSKSGSVRIWDTTAGRCERILTGHQ 247  
 QY 245 AVTAVRWGGAGLIYTSKSDRTVKMRAADGILCTFSGHAHVNNIALSTDYVLRTPFH 304  
 Db 248 SVTCLRWGGDGLLYSASODRTIKVRAHDGVLCTRLQGHGHVWNTMALSTDYALRTGAF 307  
 QY 305 PVK-DRSKRHLSTLEELQESALKRYQAVCPDEVELSVSCDDNTLYLWNNON-KCVER 362  
 Db 308 PAEASVNPQDLOGSLQELKERALSRYNLVRGQPERLYSGSDDTFLFLWSPAEDKKPLTR 367  
 QY 363 MTGHQNVNDVYKSPDKSVRLMRASDFKSVRLMRASDQGYMATFRGHVQAVYTVAMSA 422  
 Db 368 MTGHQALINQVAFSPDSRIVASAFDSFKIKLDGRTGKYLASLRGHVAVYQIAMSADR 427  
 QY 423 LIVSGSKDSTLKVNSVQTKKLAQELPCHADEVFGVDWAPDGSRVASGGKDKVIKWA 478  
 Db 428 LTVSGSDSTLKVNDYKQAKLMDLPCHADEVAVDWSFQGRVAVSGGDKDKLRIW 483

## RESULT 4

AA092844 standard; Protein; 485 AA.

XX AA092844;

XX 26-JUN-2001 (first entry)

XX Human protein sequence SEQ ID NO:11400.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX Homo sapiens.

XX EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-0116126.

XX 29-JUL-1999; 99JP-0248036.

XX 27-AUG-1999; 99JP-0300253.

XX 11-JAN-2000; 2000JP-0118776.

XX 02-MAY-2000; 2000JP-0183767.

XX 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602





Db 308 PAEASVNPQDLOGSLQELKERALSRYNLVRGQPERLVSGDDFTFLWSPAEDKKPLTR 367  
QY 363 MTGHQNVNDVYSPDKLIASAFKSVRLWRASDGOYMATFRGHVQAVYTVAMSADSR 422  
Db 368 MTGHQALINOVLFSPDSRIVASAFKSIKLDWGRGTYKYLASLRGHAAYQIAMSADSR 427  
QY 423 LTVSGSKDSTLKVMSVQTKLAQELPCHADEVFGVDNAPDGSRVASGGKDKVILW 478  
Db 428 LTVSGSDSTLKVNDYKAQKLANDLPCHADEVYAVDMSPDGQVRVASGGKDKCLRIW 483

RESULT 6  
ABB97306  
ID ABB97306 standard; Protein; 485 AA.  
XX  
AC ABB97306;  
XX  
DT 27-JUN-2002 (first entry)  
XX  
DE Novel human protein SEQ ID NO: 574.  
XX  
KW Human; antianaemic; vulnery; antinflammatory; immunomodulator;  
KW antifertility; cerebroprotective; cytostatic; rheumatic; gene therapy;  
KW neuroprotective; antiparkinsonian; protein therapy; EST;  
KW expressed sequence tag.  
XX  
OS Homo sapiens.  
XX  
PN WO200222660-A2.  
XX  
PD 21-MAR-2002.  
XX  
PF 10-SEP-2001; 2001WO-US26015.  
XX  
PR 11-SEP-2000; 2000US-0659671.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;  
PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;  
XX  
XX WPI; 2002-292408/33.  
DR N-PSDB; AEN32492.  
XX  
XX An isolated polynucleotide for treating diseases associated with its  
XX encoded polypeptide such as cancer and multiple sclerosis -  
XX  
XX Example 2; SEQ ID NO 574; 509pp; English.  
XX  
XX The present invention provides the protein and coding sequences of 444  
XX novel human proteins. These were isolated from expressed sequences tags  
XX (ESTs). They can be used to stimulate cell growth, to regulate  
XX haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth  
XX e.g. in burn treatment, to regulate the immune system e.g. to treat  
XX multiple sclerosis, to regulate activin or inhibin e.g. to treat  
XX infertility, to regulate haemostasis or thrombolysis e.g. to treat  
XX stroke and cancer, to screen for drugs, to treat inflammatory conditions  
XX e.g. rheumatoid arthritis, and to treat nervous system disorders e.g.  
XX Parkinson's disease. The present sequence is a protein of the invention.  
XX  
SQ Sequence 485 AA;  
Query Match 59.1%; Score 1520; DB 23; Length 485;  
Best Local Similarity 58.6%; Pred. No. 1.5e-134;  
Matches 279; Conservative 73; Mismatches 118; Indels 6; Gaps 3;  
QY 9 EATHTTQARLVYGTGEA---GPPIDLPAGTTQQLGICNALLKNEATPYLFFVGED 64  
Db 8 EAVARDVQRLVQFQDEGGQLLSFPDVPDITPDRQLQVNCALLAQDEPLPLAFVDA 67  
QY 65 EIKSLSDTLASVDTENVIDIYQPAVFRVPTCTSSMPGHAEAVVSLNFPDGA 124  
Db 68 EIVSSLGKLESOAVETEKVLDIIYQPAIFRVRVTRCTSSLEGHSEAVISVAFSPGK 127

QY 125 HLASGSDTTVRLNDLTETPHFTCTGHKOWVLCVSWAPDGKRLASCKAGSIIWDPEP 184  
Db 128 YLASGSGDTTVRFMDLSTETPHFTCKGHRHWLSTWSWDGKRLASCKNGQLLWDPEP 187  
QY 185 GQKGRPLSGHKHKNCLAWEPYHRDPECKRLASASGDGDCRINDVKLGOCNLNAGHTN 244  
Db 188 GKQVGRTLAGHSKWITGLSWEPHANPECRYVASSSKDGSYRIWDTTAGCERILTGHTQ 247  
QY 245 AVTAVRGGAGLIYTSKDRVKNWRAADGILCRFTSGHAWHANNIALSTDYVLRTPGFH 304  
Db 248 SVTCLRWGGDGLLYSASQDRTIKVRAHOGVLCRTLQGHGHWNTMALSTDYALRTGAFE 307  
QY 305 PVK-DRSKSHLSLSTEEQESALKRYOAVCPDEVESLYSCSDNTLVLRNNON-KCYER 362  
Db 308 PAEASVNPQDLOGSLQELKERALSRYNLVRGQPERLVSGDDFTFLWSPAEDKKPLTR 367  
QY 363 MTGHQNVNDVYSPDKLIASAFKSVRLWRASDGOYMATFRGHVQAVYTVAMSADSR 422  
Db 368 MTGHQALINOVLFSPDSRIVASAFKSIKLDWGRGTYKYLASLRGHAAYQIAMSADSR 427  
QY 423 LTVSGSKDSTLKVMSVQTKLAQELPCHADEVFGVDNAPDGSRVASGGKDKVILW 478  
Db 428 LTVSGSDSTLKVNDYKAQKLANDLPCHADEVYAVDMSPDGQVRVASGGKDKCLRIW 483

RESULT 7  
AAB68282  
ID AAB68282 standard; Protein; 484 AA.  
XX  
AC AAB68282;  
XX  
DT 09-JUL-2001 (first entry)  
XX  
DE Amino acid sequence of human Tzap gene cDNA clone Tzap7/B.  
XX  
KW Tzap; T cell activation; immune response; transplant rejection;  
KW bone marrow transplantation; rheumatoid arthritis; lupus erythematosus;  
KW multiple sclerosis; encephalitis; vasculitis; diabetes mellitus;  
KW pancreatitis; gastritis; thyroiditis; malignant disorder; asthma;  
KW lepraematosis; gastritis; skin tumour; adrenal tumour; lung tumour;  
KW wound healing; growth disorder; inflammatory disease; infectious disease.  
XX  
OS Homo sapiens.  
XX  
PN WO200132614-A2.  
XX  
PD 10-MAY-2001.  
XX  
PF 30-OCT-2000; 2000WO-EP10670.  
XX  
PR 01-NOV-1999; 99US-0162675.  
XX 23-FEB-2000; 2000US-0185016.  
XX (UTKU/) UTKU N.  
XX Utku N;  
XX WPI; 2001-316410/33.  
DR N-PSDB; AAF85073.  
XX  
XX Novel polynucleotide encoding Tzap protein involved in T cell  
XX activation, useful for diagnosing and treating diseases involving T  
XX cell activation, for treating organ transplantation rejection,  
XX rheumatoid arthritis -  
XX  
XX Example; Page 49-51; 68pp; English.  
XX  
XX The present sequence represents a Tzap protein. The Tzap gene is  
XX involved in T cell activation, and in the modulation of immune  
XX responses. Tzap polynucleotides and polypeptides are useful for  
XX diagnosing or treating acute and chronic diseases involving T cell  
XX activation and Th1 and Th2 immune response, for the treatment of acute



```
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144864.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 22-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145818.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.

PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159325.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 23-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 51.9%; Score 1335.5; DB 21; Length 471;
Best Local Similarity 52.5%; Pred. No. 3.9e-117;
Matches 256; Conservative 73; Mismatches 130; Indels 29; Gaps 6;

QY 1 MQETD-----TEQEAPTHIOARLYVTGEEAGPPIDLPAGITQOLGICNALKN 51
Db 1 MMDTDEKTVMLCLTDPEGT-HLGSA--MYIPQAGP-----LQTLQVNRFLDN 47
QY 52 EATPYLFFVGEDEIKSLEDLTLASVDVTENVIVQVQAVKVPVTRCTSSMPGHA 111
Db 48 EEMPLYSFYVDELLVPVGVYLEKNKVSVEKVLIVYQQAVERIPVNRCSQTIAGHA 107
QY 112 EAVVSLNFSFDGAHLASGDTTVRLWDLNTEPHFTCTGHKWVLCVSWAPDGKRLASG 171
Db 108 EAVLCVSFSDGKQLASGDTTVRLWDLNTEPHFTCTGHKWVLCVSWAPDGKRLASG 167
QY 172 CKAGSIIWDPETGQKGRPLSGHKKHINCLAWEPHYRDPCKRLASGSDGDCRIWDVK 231
Db 168 SRSGEICCWNPKRKEGEGSPLTGKKWITGISEVPHVLSPPCRFTVTSKDGDAIWDIT 227
QY 232 LGQCLMNIAGHTNAVTRWGGAGLIYTSKDRITYKMMRAADGTLCTFTSGHAHVNNIA 291
Db 228 LKKSIIICLSGHTLAVTCVKMGDGIYTGSDQCTTKMWTGQKLIRELKRGHWINSIA 287
QY 292 LSTDYVLRTPPHVPKDRSKSHLSLSTELQBSALKRYQAVCPDEVESLVSCSDNTLYL 351
Db 288 LSTEVVLRGAF-----DHTGRQYPPNEEKQKALERTKNGDSPERLYSGDDFTWFL 341
QY 352 WRNNONK-CVERMTGHONVYNDKYSPDKLITASASDKSVLRWRASDGOYMATRGHVQ 410
Db 342 WEFSVSKQPKRLTGHOQLVNHVYFSPDGKWTASASDKSVLRWNGITQGFVYVTRGVG 401
QY 411 AVYTVASADRLIVSGSKDSTLKVWSQTKKLAQEPGHADYFVGVDMADPGSRVASGG 470
Db 402 PYQVSWASADRLIVSGSKDSTLKVWSQTKKLAQEPGHADYFVGVDMADPGSRVASGG 461
QY 471 KDKVVKLW 478
Db 462 KDRVLKWL 469
```

RESULT 9  
AAG46521  
ID AAG46521 standard; Protein; 471 AA.  
XX AC AAG46521;  
XX DT 18-OCT-2000 (first entry)  
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 58536.  
XX KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX OS Arabidopsis thaliana.  
XX PN BP1033405-A2.  
XX PD 06-SEP-2000.  
XX PF 25-FEB-2000; 2000EP-0301439.  
XX PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 03-MAR-1999; 99US-0123548.  
PR 25-MAR-1999; 99US-0125788.  
PR 29-MAR-1999; 99US-0126264.  
PR 01-APR-1999; 99US-0126785.  
PR 06-APR-1999; 99US-0127462.  
PR 08-APR-1999; 99US-0128234.  
PR 16-APR-1999; 99US-0128714.  
PR 19-APR-1999; 99US-0129845.  
PR 21-APR-1999; 99US-0130077.  
PR 23-APR-1999; 99US-0130449.  
PR 28-APR-1999; 99US-0130510.  
PR 30-APR-1999; 99US-0130691.  
PR 04-MAY-1999; 99US-0131449.  
PR 05-MAY-1999; 99US-0132048.  
PR 06-MAY-1999; 99US-0132407.  
PR 07-MAY-1999; 99US-0132484.  
PR 11-MAY-1999; 99US-0132485.  
PR 14-MAY-1999; 99US-0132486.  
PR 14-MAY-1999; 99US-0132487.  
PR 14-MAY-1999; 99US-0132663.  
PR 14-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 18-MAY-1999; 99US-0134370.  
PR 19-MAY-1999; 99US-0134768.  
PR 20-MAY-1999; 99US-0134941.  
PR 21-MAY-1999; 99US-0135124.  
PR 24-MAY-1999; 99US-0135353.  
PR 25-MAY-1999; 99US-0135629.  
PR 27-MAY-1999; 99US-0136021.  
PR 28-MAY-1999; 99US-0136392.  
PR 01-JUN-1999; 99US-0136782.  
PR 03-JUN-1999; 99US-0137222.  
PR 04-JUN-1999; 99US-0137528.  
PR 07-JUN-1999; 99US-0137502.  
PR 08-JUN-1999; 99US-0137724.  
PR 10-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 14-JUN-1999; 99US-0138847.  
PR 16-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139453.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 28-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 02-JUL-1999; 99US-0142154.  
PR 06-JUL-1999; 99US-0142055.  
PR 08-JUL-1999; 99US-0142390.  
PR 09-JUL-1999; 99US-0142803.  
PR 12-JUL-1999; 99US-0142920.  
PR 13-JUL-1999; 99US-0142977.  
PR 14-JUL-1999; 99US-0143542.  
PR 15-JUL-1999; 99US-0143624.  
PR 16-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 19-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 26-JUL-1999; 99US-0145224.  
PR 27-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 28-JUL-1999; 99US-0145919.  
PR 02-AUG-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148584.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.

PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152263.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 51.9%; Score 1335.5; DB 21; Length 471;  
Best Local Similarity 52.5%; Pred. No. 3.9e-117;  
Matches 256; Conservative 73; Mismatches 130; Indels 29; Gaps 6;

QY 1 MOETD-----TEQETPHTIOARLYVTGEAGPPIDLPAGITTOQLGLICNALLKN 51  
DB 1 MNDTDSKTVNCLLTDPEGT-HLGSA--MYIFQKAGP-----LQTLQVNRFLDN 47  
QY 52 EATPFLFVGEDEIKKSLDLETDLASVDENVIDVYQOAVKVPVTRCTSNMPOHA 111  
DB 48 EEMLPYSFVSDEELLVPVGTLEKNKVSKEVLFIVYQOAVFRIRPNRCSQFIAGHA 107  
QY 112 EAVVSNLNSFDGAHLASGDDTVRLWDLNLTETPHTCTGKHQWLCVSWAPDGKRLASG 171  
DB 108 EAVLCVSFSPQKQLASGSGDTVRLWDLNLTETPHTCTGKHQWLVTAWSPDGKHLVSG 167

QY 172 CRAGSIITWDPETGQOKRPLSGHKKHINCLAWPPYHRDPECKKLASAGDGCRIWDVK 231  
DB 168 SKSGEICCNWPKKGELEGSLTGHKKWITGISWEPVHLSSPCRRFVTSKGDARIWDIT 227  
QY 232 LGQCLMNTAGHTNAVTAVRWGGAGLIYSSKDRVWKWRAADGILICRTFSGHAHWNNIA 291  
DB 228 LKKSIIICLSGHTLAVTCYKWWGGDIYTCSDCTIKMWTQTKLRELKKGHWINSLA 287  
QY 292 LSTDYLVRTGPFHPVKDRSKSHLSLSTELQESALKRYQAVCPDVESELVSCDDNTLYL 351  
DB 288 LSTEYVLRGTAF-----DHTGRQYPNEERQKALERYNKTGDSRVLVSGSDDFTMFL 341  
QY 352 WRNNQNK-CVERMTGHQNVNDVYSPDKLIASASFDKSVRLWRASDGOYMATFRGHVQ 410  
DB 342 WEPSVSKQPKKRLTGHQQLVNHVYFSPDGKWIASASFDKSVRLWNGITGQFVTVFRGHV 401  
QY 411 AVYTVAWSADSRILVSGSKDSTLKVMSVOTKKLAQELPGHADEVFCVDWAPDGSRVASGG 470  
DB 402 PYYQVWSNADSRLLSGSKDSTLKWEIRTKKLQDLPGHADEVFAVDPGDKVYVSGG 461  
QY 471 KDKVILKW 478  
DB 462 KDRVLKWLW 469

## RESULT 10

AAG23140  
ID AAG23140 standard; Protein: 473 AA.

XX AAG23140;

DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 26336.

KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.

OS Arabidopsis thaliana.

PN EP1033405-A2.

PD 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

PR 25-FEB-1999; 99US-0121825.

PR 03-MAR-1999; 99US-0123180.

PR 03-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 23-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 23-APR-1999; 99US-0130891.

PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 30-APR-1999; 99US-0132407.

PR 04-MAY-1999; 99US-0132484.

PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 06-MAY-1999; 99US-0132487.

PR 07-MAY-1999; 99US-0132863.

PR 11-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 21-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 28-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0141842.  
PR 02-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 18-JUL-1999; 99US-0144325.  
PR 18-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 21-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 27-JUL-1999; 99US-0145951.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 13-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.

```
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
PR 29-OCT-1999; 99US-0162142.

Query Match 51.98; Score 1335.5; DB 21; Length 473;
Best Local Similarity 52.98; Pred. No. 3.9e-117;
Matches 256; Conservative 73; Mismatches 130; Indels 29; Gaps 6;

QY 1 MQETD-----TEQATPHITQARLYTGEAGPPIDLPAGITTOOLGLICNALLN 51
DQ 3 MMDTDEGTVNCLTDPGT-HLGSA--MYIPQKAGP-----LQTLQVNRFLDN 49
QY 52 EATPYLFVGEDEIKKSLDLDLASVDYDNIIVYQPAVKVPVTRCTSMFPHA 111
DQ 50 EEMLPYSFVSDDELLVPVGYLKNKYVSEKVLTIYVQQAVFIRPVNRCSTIAGHA 109
QY 112 EAVVSLNFSPOGHLASGDTTVRLWDLNTVETPHFTCTGKOWVLCYSWAPDGRKLSG 171
DQ 110 EAVLCVSFSPGKOLASGDTTVRLWDLTYTETPLFTCKGKKNWLTVANSPDGKHLVSG 169
QY 172 CRAGSIITWDPETQQGRPLSGHKHINCLANPEYHRDPECKLASAGDGRINDVK 231
DQ 170 SKSGEICCNPKKGEGLTGHKKWITGTSWEPVHLSPCRRFVTSKGDARINDIT 229
QY 232 LQOCLMNIAGHTNAVAVRMGGAGLIYTSKDRTVKMRADGILCRTFSGHAHWNNIA 291
DQ 230 LAKSIICLSGHTLAVTCVKGWGDYIYTGSDCTIKMWTQCKLIRELKGHHWNSLA 289
QY 292 LSTDVVLRTGPFHPVKDRSKSHLSLSTELQESALKRYQAVCPDESVESVCSDDNTLYL 351
DQ 290 LSTEIVLRTGAF-----DHTGQYPPNEXQKALERYNKTGDSERLVSGDDFTFEL 343
QY 352 WRNNQNK-CVERMTGHONVNDVKYSPDVKLIASAFDKSVRLWRASDQYMATFRGHVQ 410
DQ 344 WEPFSVKQPKRLTGHQVLNVHVFSPDGKWIASAFDKSVRLWNGITGQFVYFRGHVG 403
QY 411 AVYTVAWSADSRILVSGSDTLKVNVSQTKKLAQELPHGADEVFQVWADPGSRVASG 470
DQ 404 PYQVSWASADSRILVSGSDTLKWIETRKQLKQDLPHGADEVFQVWADPGSRVASG 463
QY 471 KQVVKLW 478
DQ 464 KQVVKLW 471

RESULT 11
AAG46520
ID AAG46520 standard; Protein; 473 AA.
XX
AC AAG46520;
XX
DT 18-OCT-2000 (first entry)
DE
DE Arabidopsis thaliana protein fragment SEQ ID NO: 58535.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PF 25-FEB-1999; 99US-0121825.
XX
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0128845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140825.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
```





RESULT 12  
AAG23139  
ID AAG23139 standard; Protein; 490 AA.  
XX AC AAG23139;  
XX DT 17-OCT-2000 (first entry)  
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 26335.  
XX KW Protein identification; signal transduction pathway; metabolic pathway;  
XX KW hybridisation assay; genetic mapping; gene expression control; promoter;  
XX KW termination sequence.  
XX OS Arabidopsis thaliana.  
XX PN EP1033405-A2.  
XX PD 06-SEP-2000.  
XX PF 25-FEB-2000; 2000EP-0301439.  
XX PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130445.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145216.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 05-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 06-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 09-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.

```

PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 28-OCT-1999; 99US-0162142.

Query Match 51.9%; Score 1335.5; DB 21; Length 490;
Best Local Similarity 52.5%; Pred. No. 4.1e-117;
Matches 256; Conservative 73; Mismatches 130; Indels 29; Gaps 6;

Qy 1 MOETD-----TEQEAHTHTQARLVYTGEEAGPPIDLPAGITQQGLICNALLKN 51
    | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 20 MMDTDEKTVMLCLTDPECT-HLGS--MVIPOKAGP-----LQTLQLVNRFDN 66
    | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 52 EEAFTPLFFVGEDEIKKSLDLDASVDENVIDIVQOAVKVPVTRCTSSMPGHA 111
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 67 EEMLPYEFYDSEELVPVGYIENKVSVEKVLTVIQQAVFIRVNRCSQTAGHA 126
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 112 EAVVSLNFPSPDGAHLASGSDFTVRLMDLNTETPHFTCTGKHQWVLCVSNAPDGKRLAG 171
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

Db 127 EAVLCVSFSPDGKQOLASGSDTTVRLMDLYTETPLFTCKGKHKNWLVTVAWSPDGKHLVSG 186
Qy 172 CKAGSIIWDPETGOQGRPLSGHKHINCLAWPFYHRDPECKKLASASGSDGCRINDVK 231
    | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 187 SKSGETCCNPNKKEGLESPLTGHKKWITGSWEPEVHLSSPCRRFVTSSKGDARINDIT 246
    | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 232 LGQCLMNIAGHTNAVTAVRWGGAGLIYTSKDRIVKWRADGILCARTFSGHAHVNINIA 291
    | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 247 LKKSIIICLSGHTLAVTCVKWGGDIIYTGSDQCTIKWETTQCKLIRELKGHGHWINSLA 306
    | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 292 LSTDYVLRTGPFHPVKDRSKSHLSLSTELQESALKRYQAVCPDEVEVSLVSGSDNTLYL 351
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 307 LSTEIVLRIGAF-----DHIGQYPPNEEKQKALERYNKTGDSPELVYSGDDFTMFL 360
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 352 WRNNQNK-CVERMTGHONVYNDVKYSPDVKLIASASFDKSVRLWRASDQGYMATFRGHVQ 410
    | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 361 WEPVSVKQPKKRLTGHQOLVNHVYFSPDGKWTIASASFDKSVRLWNGITQGFVTVFRGHV 420
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 411 AVYTVAMSADSRLLVSGSKDSTLKVMSVQTKLAQELFGHADEVFGVDWAPDGSFVASGG 470
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 421 PVTQVWSADSRLLSGSKDSTLKIWEIRTKLQDLFGHADEVFAVDWSPDGEKVVSGG 480
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 471 KDKVIKLV 478
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 481 KDRVLKLV 488
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 13
AAG46519
ID AAG46519 standard; Protein; 490 AA.
XX
AC AAG46519;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 58534.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EPI033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.

```

PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 21-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139452.  
PR 18-JUN-1999; 99US-0139452.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139889.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143824.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144864.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0158458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160880.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 22-OCT-1999; 99US-0160989.

```
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 26-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 51.9%; Score 1335.5; DB 21; Length 490;
Best Local Similarity 52.5%; Pred. No. 4.1e-117;
Matches 256; Conservative 73; Mismatches 130; Indels 29; Gaps 6;

Qy 1 MQETD-----TEQEAHTHTQARLVYTGEEAGPPIDLPAGITTOQLGLICNALLKN 51
Db 20 MMDTDEKTVNCLLTDPEGT-HLGSA--MYIPQKAGP-----LQLTQLVNRFLDN 66
Qy 52 EEAATPYLFFVGEDEIKKSLDLDIASVDYDENVIDIVYQPAVFKVRPVRTCTSSMPGHA 111
Db 67 EEMLPYSFYVSDELLVPVGYLTKNKSVEKVLITVYQQQAVFIRPVNRKCSQTIAHGA 126
Qy 112 EAVYSLNFSPDGAHLASGSDTIVRLWDLNTEPHTCTGKHQWVLCVSWAPDGKRLASG 171
Db 127 EAVLCVSPDGLKASGSDTIVRLWDLNTEPHTCTGKHQWVLCVSWAPDGKRLASG 186
Qy 172 CKAGSIIWDPETGQOKGRPLSGHKHINCLAWEPYHRDPECKRLASGSDGDCRIWDYK 231
Db 187 SKSGEICCNWPKKGELEGLPLTGKHKWITGISWEPVHLSPPCRVTSKDGARLWDT 246
Qy 232 LQCLMINTAGHTNAVTVAVWGGAGLIYTSKDRYVKNMRAADGILCFTFGSHAHWNNTA 291
Db 247 LKKSICLSGHTLAVTCVWGGDGIYITGSDCTIKMWTGQKILRLKGGHGWINSIA 306
Qy 292 LSTDYVLRTPGTFPHVVKDRSKSHLSLSTEELESALKRYQAVCPDEVESILVSCSDNTLYL 351
Db 307 LSTEVLRGTAF-----DHTGRQYPPNEEKQKALERYNKTGDSPERLYSGSDDTMFL 360
Qy 352 WENONK-CYERWTHQNVVNDVKYSPDVKLIASAFDKSVRLWRASDGYMATFGRHVQ 410
Db 361 WEPVSQKPKRLTGHOQLVNVHYFSPDGKWIASAFDKSVRLWNGITGQEVTVFGRHV 420
Qy 411 AVYTVANSADSRILVSGKSDTLKVMVSVQTKKLAQELPGHADEVFGVDWAPDGSRVASG 470
Db 421 PVYQVSNASDRLLSLSGSKDSTLKIKWEIRTKKQLQDLPGHADEVFAVDWSPDGK 480
Qy 471 KDKVIKWL 478
Db 481 KDRVLKLW 488

RESULT 14
AAR85881
ID AAR85881 standard; Protein; 514 AA.
XX
AC AAR85881;
XX
DT 13-SEP-1996 (first entry)
XX
DE WD-40 domain-contg. YCW2 protein.
XX
KW WD40 repeat region; beta-transducin; protein-protein interaction; drug;
KW intracellular signalling; protein kinase C; homology; motif; modulator;
KW receptors of activated protein kinase; enzyme activity; isozyme; human.
XX
OS Synthetic.
XX
PN W09521252-A2.
XX
PD 10-AUG-1995.
XX
PF 31-JAN-1995; 95WO-US01210.
```

```
XX 01-FEB-1994; 94US-0190802.
PR
XX (STRD ) UNIV LELAND STANFORD JUNIOR.
XX
XX Mochly-Rosen D, Ron D;
XX
XX WPI; 1995-283772/37.
XX
XX New WD-40 (beta-transducin)-derived polypeptide(s) - which alter the
XX activity of a protein, eg. protein kinase C, which interacts with a
XX protein contg. a WD-40 region.
XX
XX Example 5; Page 168-170; 351pp; English.
XX
XX Proteins AAR85851-92 are protein which contain at least one WD-40 (also
XX called beta-transducin homologous) amino acid repeat motifs. The WD-40
XX regions are involved in protein-protein interactions between proteins
XX involved in intracellular signalling. An example of such an interaction
XX is between protein kinase C and receptors of activated protein kinase
XX (RACK), esp. RACK-1 (AAR85850). Proteins AAR85851-82 were isolated
XX on homology with beta-transducin, whereas proteins AAR85882-92 were
XX isolated based on homology with the WD-40 consensus sequence (AAR85893).
XX The proteins were used to construct the peptides AAR84928-R85063 and
XX AAR85786-R85842. The peptides can be used to identify target proteins
XX contg. WD-40 motifs, as modulators of enzyme esp. isozyme, activity of
XX proteins involved in protein-protein interaction and to screen for drugs
XX that will affect protein-protein interaction involving WD-40 domains.
XX
XX Sequence 514 AA;

Query Match 41.1%; Score 1058.5; DB 16; Length 514;
Best Local Similarity 42.5%; Pred. No. 6.6e-91;
Matches 210; Conservative 97; Mismatches 148; Indels 39; Gaps 12;

Qy 14 TIQARLVYTGEEAGPPIDLPAGITTOQLGLICNALL-KNEEAETPYLFFV-----GEDEI 66
Db 33 SIRFOALDTGDNVGGALRPGAISEKQLELLNQLNGTSDDPVPYPTFSCITOCKKASDPV 92
Qy 67 KSLSDTLDLAS-----VDENVIDIVYQPAVFKVRPVRTCTSSMPGHAEAVV 115
Db 93 K-----TIDITNLYSLIKPGYNSTEDQITLLYTPRAVFKVPVTRSSAIAHGSTIL 147
Qy 116 SLNFSF-DGAHLASGSDTIVRLWDLNTEPHTCTGKHQWVLCVSWAPDGKRLASGCKA 174
Db 148 CSAFAPHTSSRVTVAGDNTARWDCDQTPMHTLKGHNWVLCVSWPDGEVIATGMD 207
Qy 175 GSIIWDPETGQOKGRPLSGHKHINCLAWEPYH--RDPECKRLASGSDGDCRIWDYK 232
Db 208 NTIRLWDPKSGQCLGDLRGHKSWITSLSWEPHILVKFGSKPRLASSKDGTIKIDWTVS 267
Qy 233 GQCLMINTAGHTNAVTVAVWGGAGLIYTSKDRYVKNMRA-ADGILCFTFGSHAHWNNTA 291
Db 268 RVQYTMSTGTSVSCVKGWGGQGLYSGSHDRTVRVWDINSOGRICNILKSHAHWNHLS 327
Qy 292 LSTDYVLRTPGTF-HPVKDRSKSHLSLSTEELESALKRYQAVCP--DEVESILVSCSDN 347
Db 328 LSTDYALRIGAFDHTGKPS-----TPEAOKKALENYEKICKNKNSEEMVYASDDY 381
Qy 348 TLYLWRN-NONKCYERWTHQNVVNDVKYSPDVKLIASAFDKSVRLWRASDGYMATFTR 406
Db 382 TMFLWNLPLKSTKPIARMTGHOKLVNHFVSPDGRYIVSASFDSIKILWDRGRGKFTSTFR 441
Qy 407 GHVQAVTVYVANSADSRILVSGKSDTLKVMVSVQTKKLAQELPGHADEVFGVDWAPDGSRV 466
Db 442 GHIASVYQVANSDCRLVSCSDTTLKVMVDRVTRKLSVLDLPGLIKTKLY-VQMSVDGKRV 500
Qy 467 ASGGKDKVIKWLWAY 480
Db 501 CSGGKDKVRLWTH 514

RESULT 15
```





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 2, 2003, 06:58:20 ; Search time 35 Seconds  
(without alignments)  
403.514 Million cell updates/sec

Title: US-09-830-980-1  
Perfect score: 2574  
Sequence: 1 MQEYDTEQATPHITQARLV.....PDGSRVSSGDKVKIKLWAY 480

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5A-COMB.pep:\*  
2: /cgn2\_6/ptodata/1/iaa/5B-COMB.pep:\*  
3: /cgn2\_6/ptodata/1/iaa/6A-COMB.pep:\*  
4: /cgn2\_6/ptodata/1/iaa/6B-COMB.pep:\*  
5: /cgn2\_6/ptodata/1/iaa/PCFUS-COMB.pep:\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1058.5	41.1	514	1 US-08-190-802A-66	Sequence 66, Appl
2	1058.5	41.1	514	4 US-08-477-346-66	Sequence 66, Appl
3	1058.5	41.1	514	4 US-08-473-089-66	Sequence 66, Appl
4	1058.5	41.1	514	4 US-08-487-072A-66	Sequence 66, Appl
5	383.5	14.9	251	4 US-09-291-170A-13	Sequence 13, Appl
6	383.5	14.9	251	4 US-09-724-884-13	Sequence 13, Appl
7	354	13.8	409	2 US-08-283-917-3	Sequence 3, Appl
8	354	13.8	409	2 US-08-961-716-3	Sequence 3, Appl
9	354	13.8	410	2 US-08-283-917-9	Sequence 9, Appl
10	354	13.8	410	2 US-08-961-716-9	Sequence 9, Appl
11	340.5	13.2	409	1 US-08-190-802A-51	Sequence 51, Appl
12	340.5	13.2	409	4 US-08-477-346-51	Sequence 51, Appl
13	340.5	13.2	409	4 US-08-473-089-51	Sequence 51, Appl
14	340.5	13.2	409	4 US-08-487-072A-51	Sequence 51, Appl
15	323.5	12.6	343	4 US-09-063-743-5	Sequence 5, Appl
16	323.5	12.6	343	4 US-09-590-540-5	Sequence 5, Appl
17	312	12.1	1194	4 US-09-092-508-2	Sequence 2, Appl
18	312	12.1	1194	4 US-09-435-115-2	Sequence 2, Appl
19	312	12.1	1194	4 US-09-069-023-26	Sequence 26, Appl
20	312	12.1	1194	4 US-09-098-310-2	Sequence 2, Appl
21	312	12.1	1205	4 US-09-092-508-16	Sequence 16, Appl
22	312	12.1	1205	4 US-09-435-115-16	Sequence 16, Appl
23	310.5	12.1	640	4 US-09-177-165A-30	Sequence 30, Appl
24	308	12.0	375	4 US-09-063-743-1	Sequence 1, Appl
25	308	12.0	375	4 US-09-590-540-1	Sequence 1, Appl
26	306	11.9	517	1 US-08-190-802A-30	Sequence 30, Appl
27	306	11.9	517	4 US-08-477-346-30	Sequence 30, Appl

28	306	11.9	517	4	US-08-473-089-30	Sequence 30, Appl
29	306	11.9	517	4	US-08-487-072A-30	Sequence 30, Appl
30	305	11.8	514	4	US-09-108-857-2	Sequence 2, Appl
31	302.5	11.8	2627	2	US-08-751-189-3	Sequence 3, Appl
32	302.5	11.8	2627	2	US-09-060-836-3	Sequence 3, Appl
33	302.5	11.8	2627	4	US-09-184-445-3	Sequence 3, Appl
34	297	11.5	587	3	US-08-899-578-2	Sequence 2, Appl
35	294.5	11.4	713	1	US-08-190-802A-63	Sequence 63, Appl
36	294.5	11.4	713	4	US-08-477-346-63	Sequence 63, Appl
37	294.5	11.4	713	4	US-08-473-089-63	Sequence 63, Appl
38	294.5	11.4	713	4	US-08-487-072A-63	Sequence 63, Appl
39	294.5	11.4	713	4	US-09-108-857-3	Sequence 3, Appl
40	294	11.4	253	4	US-09-291-170A-10	Sequence 10, Appl
41	294	11.4	253	4	US-09-724-884-10	Sequence 10, Appl
42	294	11.4	690	4	US-09-291-170A-2	Sequence 2, Appl
43	294	11.4	690	4	US-09-724-884-2	Sequence 2, Appl
44	292.5	11.4	404	4	US-09-184-001-2	Sequence 2, Appl
45	291	11.3	704	1	US-08-188-582-5	Sequence 5, Appl

## ALIGNMENTS

RESULT 1  
US-08-190-802A-66  
; Sequence 66, Application US/08190802A  
; Patent No. 5519003  
; GENERAL INFORMATION:  
; APPLICANT: Mochly-Rosen, Daria  
; TITLE OF INVENTION: WO-40 - Derived Peptides and Uses  
; NUMBER OF INVENTIONS: Thereof  
; NUMBER OF SEQUENCES: 265  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dellinger & Associates  
; STREET: P.O. Box 60850  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94306-0850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/190,802A  
; FILING DATE: 01-FEB-1994  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fabian Gary R. 33,875  
; REGISTRATION NUMBER: 8600-0139  
; REFERENCE/DOCKET NUMBER: 324-0880  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 324-0880  
; TELEFAX: (415) 324-0960  
; INFORMATION FOR SEQ ID NO: 66:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 514 amino acids  
; TYPE: amino acid  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; INDIVIDUAL ISOLATE: YCW2 PROTEIN, Fig. 49  
US-08-190-802A-66

Query Match 41.1%; Score 1058.5; DB 1; Length 514;  
Best Local Similarity 42.5%; Pred. No. 8.5e-96;  
Matches 210; Conservative 97; Mismatches 148; Indels 39; Gaps 12;  
QY 14 TIOARLYVTGEAGPPIIDLPAGITTOQLGLICNALL-KNEEATPYLFFV-----GDEI 66

Db 33 SIKFOALDTGDNVGGALRVPGAISEKQLELLNQLNGTSDDPVPTFSCITQKKASDPV 92  
Qy 67 KKSLEDTLDLAS-----VDTENVIDIVYQAVFKVRPVTRCTSSMPGHAFAV 115  
Db 93 K-----TIDITDNLSSLIKPGYNSTEDQITLLYTPRAVFKVPTRESSAAGHGSIL 147  
Qy 116 SLNESP-DGAHLASGSDTTVRLWDLNTEPHFTCTGKHQWVLCVSWAPDGRKASGCKA 174  
Db 148 CSAFAPHTSSRMVGTAGDNTARINDCTQPMHTLKGHNWVLCVSWSPDGEVIATGSM 207  
Qy 175 GSIIWDPETGQKGRPLSGHKKHINCLAWEPYH--RDEPCRKLASAGDGDGRINDVKL 232  
Db 208 NTRILWPKSGQGLDGLRHSKWITSLSWEPHLVKGPKSPRLASSKDGITIKWDTVS 267  
Qy 233 GQCLMNIAGHTNAVTAVRWGGAGLIYTSKDRITVKMRA--ADGILCFTSGHAHVNNTA 291  
Db 268 RVQYTMSTGHTNSVSCVKGWGGGLLYSGSHDRTVRVMDINSQGRGINILKSHAHVNHLS 327  
Qy 292 LSTDYVLTGPF--HPVKDRSKSHLSLSTEELQESALKRYQAVCP---DEVESLVSCDDN 347  
Db 328 LSTDYALRIGAFDHTGKPS-----TPBEAQKALENVEKICKKNGSEMMVTASDDY 381  
Qy 348 TLYLWRN-NQNKVERMTGHONVNDYKSPDVKLIASAFDKSVRLWRASDGYMATER 406  
Db 382 TFLNPLKSTKPIARMTGHOKLVNHFVSPDGRYIVASFDNSIKLWDGRDGFISTFR 441  
Qy 407 GHVQAVTYVAMSADSLIVSGSKDSTLKVWSVOTKKLAQELPGHADEVFGVDWAPDGRV 466  
Db 442 GHIASVYQVAMSSDCRLVSCSKDTTLKVDVTRKLSVLDLPGLKIKLY--VDMVSDGKRV 500  
Qy 467 ASGGKRVIKIMAY 480  
Db 501 CSGGKDKMVLWTH 514

## RESULT 2

US-08-477-346-66  
; Sequence 66, Application US/08477346  
; Patent No. 6262023  
; GENERAL INFORMATION:  
; APPLICANT: Mochly-Rosen, Daria  
; APPLICANT: Ron, Dorit  
; TITLE OF INVENTION: WD-40 - Derived Peptides and Uses  
; TITLE OF INVENTION: Thereof  
; NUMBER OF SEQUENCES: 265  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Morrison & Foerster  
; STREET: 2000 Pennsylvania Avenue, NW  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20006-1812  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/477,346  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/487,072  
; FILING DATE: 07-JUN-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MURASHIGE, KATE H.  
; REGISTRATION NUMBER: 29,959  
; REFERENCE/DOCKET NUMBER: 2550-0025.20  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 887-1500  
; TELEFAX: (202) 887-0763  
; INFORMATION FOR SEQ ID NO: 66:

; SEQUENCE CHARACTERISTICS:  
; LENGTH: 514 amino acids  
; TYPE: amino acid  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
; HYPOTHEICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; INDIVIDUAL ISOLATE: YCW2 PROTEIN, Fig. 49  
US-08-477-346-66

Query Match 41.1%; Score 1058.5; DB 4; Length 514;  
Best Local Similarity 42.5%; Pred. No. 8.5e-98;  
Matches 210; Conservative 97; Mismatches 148; Indels 39; Gaps 12;  
Qy 14 TIQARLYVTGEEAGPPIDLPAGITITQGLICNALI--KNEEATPYLFV-----GEDRI 66  
Db 33 SIKFOALDTGDNVGGALRVPGAISEKQLELLNQLNGTSDDPVPTFSCITQKKASDPV 92  
Qy 67 KKSLEDTLDLAS-----VDTENVIDIVYQAVFKVRPVTRCTSSMPGHAFAV 115  
Db 93 K-----TIDITDNLSSLIKPGYNSTEDQITLLYTPRAVFKVPTRESSAAGHGSIL 147  
Qy 116 SLNESP-DGAHLASGSDTTVRLWDLNTEPHFTCTGKHQWVLCVSWAPDGRKASGCKA 174  
Db 148 CSAFAPHTSSRMVGTAGDNTARINDCTQPMHTLKGHNWVLCVSWSPDGEVIATGSM 207  
Qy 175 GSIIWDPETGQKGRPLSGHKKHINCLAWEPYH--RDEPCRKLASAGDGDGRINDVKL 232  
Db 208 NTRILWPKSGQGLDGLRHSKWITSLSWEPHLVKGPKSPRLASSKDGITIKWDTVS 267  
Qy 233 GQCLMNIAGHTNAVTAVRWGGAGLIYTSKDRITVKMRA--ADGILCFTSGHAHVNNTA 291  
Db 268 RVQYTMSTGHTNSVSCVKGWGGGLLYSGSHDRTVRVMDINSQGRGINILKSHAHVNHLS 327  
Qy 292 LSTDYVLTGPF--HPVKDRSKSHLSLSTEELQESALKRYQAVCP---DEVESLVSCDDN 347  
Db 328 LSTDYALRIGAFDHTGKPS-----TPBEAQKALENVEKICKKNGSEMMVTASDDY 381  
Qy 348 TLYLWRN-NQNKVERMTGHONVNDYKSPDVKLIASAFDKSVRLWRASDGYMATER 406  
Db 382 TFLNPLKSTKPIARMTGHOKLVNHFVSPDGRYIVASFDNSIKLWDGRDGFISTFR 441  
Qy 407 GHVQAVTYVAMSADSLIVSGSKDSTLKVWSVOTKKLAQELPGHADEVFGVDWAPDGRV 466  
Db 442 GHIASVYQVAMSSDCRLVSCSKDTTLKVDVTRKLSVLDLPGLKIKLY--VDMVSDGKRV 500  
Qy 467 ASGGKRVIKIMAY 480  
Db 501 CSGGKDKMVLWTH 514

## RESULT 3

US-08-473-089-66  
; Sequence 66, Application US/08473089  
; Patent No. 6342368  
; GENERAL INFORMATION:  
; APPLICANT: Mochly-Rosen, Daria  
; APPLICANT: Ron, Dorit  
; TITLE OF INVENTION: WD-40 - Derived Peptides and Uses  
; TITLE OF INVENTION: Thereof  
; NUMBER OF SEQUENCES: 265  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Morrison & Foerster  
; STREET: 2000 Pennsylvania Avenue, NW  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20006-1812  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS









STREET: 1755 S. Jefferson Davis Highway, Suite 400  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/283,917  
FILING DATE: 03-AUG-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 209943/1993  
FILING DATE: 03-AUG-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Oblon, No. 5849557man F.  
REGISTRATION NUMBER: 24,618  
REFERENCE/DOCKET NUMBER: 2292-030-0  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 413-3000  
TELEFAX: (703) 413-2220  
TELEX: 248855 OPAT UR  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 410 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-283-917-9

Query Match 13.8%; Score 354; DB 2; Length 410;  
Best Local Similarity 24.1%; Pred. No. 4.5e-27;  
Matches 85; Conservative 65; Mismatches 135; Indels 68; Gaps 5;  
QY 147 FCTGTHKQWVLCVSWAPDGRKLAGCKAGSIIDWPTGQOQGRPLSGHKKHINCLAWEP 206  
Db 102 YALSGHRSPTVRVIFHPVFSVMVSASEDAIKWDEYDGFPE-RTLAGHDSVQDISF-- 158  
QY 207 YHRDPECKLASASGDGCRINDVKLGQCLMNIAGHTNAVYR-WGAGLIYTSKDR 265  
Db 159 ---DHSGLKLLASCADMTIKLWDFQGFECIRTMHGDHNVSSVAIMPNGDHIVSASRDKT 215  
QY 266 VKMRAADGILCRFTSGHAHWNNIALSTDYVLRTPFPVDPKDRSKSHLSLSTEELOESA 325  
Db 216 IKMNEVQTGYCVKFTTGHREWRVMPVNPQDGLT----- 248  
QY 326 LKRYQAVCPDEVESLVSCSDNTLYLWRNNQKVCERMTGHQNVVNDKYSPDVK----- 380  
Db 249 -----IASCSNDQTVRVVWVATKECKAELREHEHVVECSWAPESYSIS 294  
QY 381 -----LIASASFDKSVLRWASDQYMATFRGHVQAVYTVANSADSLIV 425  
Db 295 EATGSETKSGKPGPFLLSGSRDKTIKMDVSTGMCMLTLVGHNDWYRGVLFHSGGKFL 354  
QY 426 SGKSDSTLKWVSTKLAQELPGHADEVFGVDWAPDGSRVASGGKDKVIKWL 478  
Db 355 SCADDKTLRVNDYKNKRCMKTINAHEFVTSDFHFKTAPYVTVGSDQTVKVV 407

RESULT 10  
US-08-961-716-9

Sequence 9, Application US/08961716  
Patent No. 5880272  
GENERAL INFORMATION:  
APPLICANT: ADACHI, HIDEKI  
APPLICANT: TSUJIMOTO, MASAFUMI  
APPLICANT: INOUE, KEIZO  
APPLICANT: ARAI, HIROYUKI  
TITLE OF INVENTION: OXIDIZED PHOPHOLIPID DEGRADING ENZYME  
TITLE OF INVENTION: AND GENE THEREOF

NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER &  
ADDRESSEE: NEUSTADT,P.C.  
STREET: 1755 S. Jefferson Davis Highway, Suite 400  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/961,716  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/283,917  
FILING DATE: 03-AUG-1994  
APPLICATION NUMBER: JP 209943/1993  
FILING DATE: 03-AUG-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Oblon, No. 5880272man F.  
REGISTRATION NUMBER: 24,618  
REFERENCE/DOCKET NUMBER: 2292-030-0  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 413-3000  
TELEFAX: (703) 413-2220  
TELEX: 248855 OPAT UR  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 410 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-961-716-9

Query Match 13.8%; Score 354; DB 2; Length 410;  
Best Local Similarity 24.1%; Pred. No. 4.5e-27;  
Matches 85; Conservative 65; Mismatches 135; Indels 68; Gaps 5;  
QY 147 FCTGTHKQWVLCVSWAPDGRKLAGCKAGSIIDWPTGQOQGRPLSGHKKHINCLAWEP 206  
Db 102 YALSGHRSPTVRVIFHPVFSVMVSASEDAIKWDEYDGFPE-RTLAGHDSVQDISF-- 158  
QY 207 YHRDPECKLASASGDGCRINDVKLGQCLMNIAGHTNAVYR-WGAGLIYTSKDR 265  
Db 159 ---DHSGLKLLASCADMTIKLWDFQGFECIRTMHGDHNVSSVAIMPNGDHIVSASRDKT 215  
QY 266 VKMRAADGILCRFTSGHAHWNNIALSTDYVLRTPFPVDPKDRSKSHLSLSTEELOESA 325  
Db 216 IKMNEVQTGYCVKFTTGHREWRVMPVNPQDGLT----- 248  
QY 326 LKRYQAVCPDEVESLVSCSDNTLYLWRNNQKVCERMTGHQNVVNDKYSPDVK----- 380  
Db 249 -----IASCSNDQTVRVVWVATKECKAELREHEHVVECSWAPESYSIS 294  
QY 381 -----LIASASFDKSVLRWASDQYMATFRGHVQAVYTVANSADSLIV 425  
Db 295 EATGSETKSGKPGPFLLSGSRDKTIKMDVSTGMCMLTLVGHNDWYRGVLFHSGGKFL 354  
QY 426 SGKSDSTLKWVSTKLAQELPGHADEVFGVDWAPDGSRVASGGKDKVIKWL 478  
Db 355 SCADDKTLRVNDYKNKRCMKTINAHEFVTSDFHFKTAPYVTVGSDQTVKVV 407

RESULT 11  
US-08-190-802A-51  
Sequence 51, Application US/08190802A  
Patent No. 5519003  
GENERAL INFORMATION:

APPLICANT: Mochly-Rosen, Daria  
APPLICANT: Ron, Dorit  
TITLE OF INVENTION: WD-40 - Derived Peptides and Uses  
TITLE OF INVENTION: Thereof  
NUMBER OF SEQUENCES: 265  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dehlinger & Associates  
STREET: P.O. Box 60850  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94306-0850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/190.802A  
FILING DATE: 01-FEB-1994  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Fabian, Gary R.  
REGISTRATION NUMBER: 33,875  
REFERENCE/DOCKET NUMBER: 8600-0139  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 324-0880  
TELEFAX: (415) 324-0960  
INFORMATION FOR SEQ ID NO: 51:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 409 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
INDIVIDUAL ISOLATE: LIS1 (human), Fig. 34  
US-08-190-802A-51

Query Match 13.2%; Score 340.5; DB 1; Length 409;  
Best Local Similarity 24.1%; Pred. No. 1e-25;  
Matches 85; Conservative 64; Mismatches 135; Indels 69; Gaps 6;  
Qy 147 FTCGKQWVLCVSWAPDCKKLAGCKAGSIIIDPETGQOKGRPLSGHKKHINCLAWEP 206  
Db 102 YALSGHRSPTVRIPIHPVFSVMVSASEDATIKWMDYETGDFE-RTLKGHTDSVQDISF-- 158  
Qy 207 YHRDPECRKLASASGDGCRINDVKLGQCLMNTAGHTNAVTAVR-WGAGLIYTSKDR 265  
Db 159 ---DHSGKLLASCSADMTIKLWDFQGFECIRTMHGHHDHNVSSVAIMPNGDHIVSASRDKT 215  
Qy 266 VKMRAADGILCRFTSGHAHWYNNIALSTDYVLTGTFPHVPKDRSKSHLSLSTEEQLQESA 325  
Db 216 IKMWEVQTGYCVKTFTHREWVRVVRPNQDGL- 248  
Qy 326 LKRYQAVCPDESVLSCSDNTLYLWRNNQKCVERTMGTGONVYNDVYKSPDVK----- 380  
Db 249 -----IASCSNDQTVRVWVATKEKAELEHEHVEECISWAPESYSSIS 294  
Qy 381 -----LIASAFDKSVLRASDGQYMATFRGHVQAVYTVWASDRSLIV 425  
Db 295 EATGETSKSGKPGPFLLSGRDKT-KMWDVSTGMCIMTLVGHDNWVRGVLPHSGGKFL 353  
Qy 426 SGSKDSTLKVMSVQTKKLAQELPGHAEVFGVDWAPDGSRVASGGKDKIKLW 478  
Db 354 SCADDKTLRVWDYKNKRCMKTLNAHEHFTVSLDFHKTAPYVTVGSDQTVKVV 406

RESULT 12

US-08-477-346-51  
; Sequence 51, Application US/08477346  
; Patent No. 6262023

GENERAL INFORMATION:  
APPLICANT: Mochly-Rosen, Daria  
APPLICANT: Ron, Dorit  
TITLE OF INVENTION: WD-40 - Derived Peptides and Uses  
TITLE OF INVENTION: Thereof  
NUMBER OF SEQUENCES: 265  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Morrison & Foerster  
STREET: 2000 Pennsylvania Avenue, NW  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20006-1812  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/477,346  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/487,072  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: MURASHIGE, KATE H.  
REGISTRATION NUMBER: 29,959  
REFERENCE/DOCKET NUMBER: 2550-0025.20  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 887-1500  
TELEFAX: (202) 887-0763  
INFORMATION FOR SEQ ID NO: 51:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 409 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
INDIVIDUAL ISOLATE: LIS1 (human), Fig. 34  
US-08-477-346-51

Query Match 13.2%; Score 340.5; DB 4; Length 409;  
Best Local Similarity 24.1%; Pred. No. 1e-25;  
Matches 85; Conservative 64; Mismatches 135; Indels 69; Gaps 6;  
Qy 147 FTCGKQWVLCVSWAPDCKKLAGCKAGSIIIDPETGQOKGRPLSGHKKHINCLAWEP 206  
Db 102 YALSGHRSPTVRIPIHPVFSVMVSASEDATIKWMDYETGDFE-RTLKGHTDSVQDISF-- 158  
Qy 207 YHRDPECRKLASASGDGCRINDVKLGQCLMNTAGHTNAVTAVR-WGAGLIYTSKDR 265  
Db 159 ---DHSGKLLASCSADMTIKLWDFQGFECIRTMHGHHDHNVSSVAIMPNGDHIVSASRDKT 215  
Qy 266 VKMRAADGILCRFTSGHAHWYNNIALSTDYVLTGTFPHVPKDRSKSHLSLSTEEQLQESA 325  
Db 216 IKMWEVQTGYCVKTFTHREWVRVVRPNQDGL- 248  
Qy 326 LKRYQAVCPDESVLSCSDNTLYLWRNNQKCVERTMGTGONVYNDVYKSPDVK----- 380  
Db 249 -----IASCSNDQTVRVWVATKEKAELEHEHVEECISWAPESYSSIS 294  
Qy 381 -----LIASAFDKSVLRASDGQYMATFRGHVQAVYTVWASDRSLIV 425  
Db 295 EATGETSKSGKPGPFLLSGRDKT-KMWDVSTGMCIMTLVGHDNWVRGVLPHSGGKFL 353  
Qy 426 SGSKDSTLKVMSVQTKKLAQELPGHAEVFGVDWAPDGSRVASGGKDKIKLW 478  
Db 354 SCADDKTLRVWDYKNKRCMKTLNAHEHFTVSLDFHKTAPYVTVGSDQTVKVV 406

## RESULT 13

US-08-473-089-51  
; Sequence 51, Application US/08473089  
; Patent No. 6342368  
; GENERAL INFORMATION:  
; APPLICANT: Mochly-Rosen, Daria  
; APPLICANT: Ron, Dorit  
; TITLE OF INVENTION: WD-40 - Derived Peptides and Uses  
; TITLE OF INVENTION: Thereof  
; NUMBER OF SEQUENCES: 265  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Morrison & Foerster  
; STREET: 2000 Pennsylvania Avenue, NW  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20006-1812  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/473,089  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MURASHIGE, KATE H.  
; REGISTRATION NUMBER: 29,959  
; REFERENCE/DOCKET NUMBER: 2550-0025.22  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 887-1500  
; TELEFAX: (202) 887-0763  
; INFORMATION FOR SEQ ID NO: 51:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 409 amino acids  
; TYPE: amino acid  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; INDIVIDUAL ISOLATE: LIS1 (human), Fig. 34  
US-08-473-089-51

Query Match 13.2%; Score 340.5; DB 4; Length 409;  
Best Local Similarity 24.1%; Pred. No. le-25;  
Matches 85; Conservative 64; Mismatches 135; Indels 69; Gaps 6;  
QY 147 FTCTGKHWLVCSWAPDGKRLASGCKAGSIIIDWPETGQOKGRPLSGHKHINCLAWEP 206  
Db 102 YALSGHRSFVTRVIFHPVFSVMVSASEDATIKWMDYETGDFE-RTLKHTDSVQDISF-- 158  
QY 207 YHRDPECKRLASAGDGDCHRIWDVKLGQCLMNIAGHTNAVTVR-WGGAGLIYTSKDRKT 265  
Db 159 ---DHSGKLLASCSADMTIKLWDFQGFECIRTMHGDHNVSSVAIMPNGDHIVSASRDKT 215  
QY 266 VKMWRADGILCRTFSGHAHWNNIALSTDYVLRGTFPHVFKDRSKSHLSLSTELQESA 325  
Db 216 IKMWEVQGYCVKFTGTHREWRVVRPNQDGLT----- 248  
QY 326 LKRYQAVCPDESVLSVCSDDNTLYLWRNNQNKCVERTGHONVNDVKYSPDKV----- 380  
Db 249 -----IASCNDQIVRVVVVATKECAELREHEHVVEICISWAPESYSIS 294  
QY 381 -----LIASAFDKSVRLWRASDGYMATFRGHVQAVTVVAMSADSLIV 425  
Db 295 EATGSETKSGKPGPFLLSGRDKT-KWMDVSTGCMCLTLVGHNDNVVRGVLFHSGGKFL 353  
QY 426 SGSKDSTLKVMSVQTKKLAQLPFGHADEVFGVDWAPDGSRVASGGKDKVIKWL 478  
Db 354 SCADDKTLRVWDYKNKCMKTLNAHEHFFVTSDFHKTAPYVVTGSDVQIVKVM 406

## RESULT 14

US-08-487-072A-51  
; Sequence 51, Application US/08487072A  
; Patent No. 6423684  
; GENERAL INFORMATION:  
; APPLICANT: Mochly-Rosen, Daria  
; APPLICANT: Ron, Dorit  
; TITLE OF INVENTION: WD-40 - Derived Peptides and Uses  
; TITLE OF INVENTION: Thereof  
; NUMBER OF SEQUENCES: 265  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Morrison & Foerster  
; STREET: 2000 Pennsylvania Avenue, NW  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20006-1812  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/487,072A  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MURASHIGE, KATE H.  
; REGISTRATION NUMBER: 29,959  
; REFERENCE/DOCKET NUMBER: 2550-0025.20  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 887-1500  
; TELEFAX: (202) 887-0763  
; INFORMATION FOR SEQ ID NO: 51:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 409 amino acids  
; TYPE: amino acid  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; INDIVIDUAL ISOLATE: LIS1 (human), Fig. 34  
US-08-487-072A-51

Query Match 13.2%; Score 340.5; DB 4; Length 409;  
Best Local Similarity 24.1%; Pred. No. le-25;  
Matches 85; Conservative 64; Mismatches 135; Indels 69; Gaps 6;  
QY 147 FTCTGKHWLVCSWAPDGKRLASGCKAGSIIIDWPETGQOKGRPLSGHKHINCLAWEP 206  
Db 102 YALSGHRSFVTRVIFHPVFSVMVSASEDATIKWMDYETGDFE-RTLKHTDSVQDISF-- 158  
QY 207 YHRDPECKRLASAGDGDCHRIWDVKLGQCLMNIAGHTNAVTVR-WGGAGLIYTSKDRKT 265  
Db 159 ---DHSGKLLASCSADMTIKLWDFQGFECIRTMHGDHNVSSVAIMPNGDHIVSASRDKT 215  
QY 266 VKMWRADGILCRTFSGHAHWNNIALSTDYVLRGTFPHVFKDRSKSHLSLSTELQESA 325  
Db 216 IKMWEVQGYCVKFTGTHREWRVVRPNQDGLT----- 248  
QY 326 LKRYQAVCPDESVLSVCSDDNTLYLWRNNQNKCVERTGHONVNDVKYSPDKV----- 380  
Db 249 -----IASCNDQIVRVVVVATKECAELREHEHVVEICISWAPESYSIS 294  
QY 381 -----LIASAFDKSVRLWRASDGYMATFRGHVQAVTVVAMSADSLIV 425  
Db 295 EATGSETKSGKPGPFLLSGRDKT-KWMDVSTGCMCLTLVGHNDNVVRGVLFHSGGKFL 353  
QY 426 SGSKDSTLKVMSVQTKKLAQLPFGHADEVFGVDWAPDGSRVASGGKDKVIKWL 478  
Db 354 SCADDKTLRVWDYKNKCMKTLNAHEHFFVTSDFHKTAPYVVTGSDVQIVKVM 406

RESULT 15  
US-09-063-743-5  
; Sequence 5, Application US/09063743  
; Patent No. 6242214  
; GENERAL INFORMATION:  
; APPLICANT: Bandman, Olga  
; APPLICANT: Lal, Preeti  
; APPLICANT: Guegler, Karl J.  
; APPLICANT: Corley, Neil C.  
; APPLICANT: Patterson, Chandra  
; TITLE OF INVENTION: HUMAN GTPASE-ASSOCIATED PROTEINS  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/063.743  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cerrone, Michael C.  
; REGISTRATION NUMBER: 39,132  
; REFERENCE/DOCKET NUMBER: PF-0508 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-855-0555  
; TELEFAX: 650-845-4166  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 343 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: GenBank  
; CLONE: 2281093  
; US-09-063-743-5

Query Match 12.6%; Score 323.5; DB 4; Length 343;  
Best Local Similarity 24.0%; Pred. NO. 3.9e-24;  
Matches 91; Conservative 54; Mismatches 139; Indels 95; Gaps 11;  
QY 107 MFGHAEAVVSLNFDGAGHLAGSGDVTYRLWDLNTETPHF-TCTGCHKWVLCVSWAPDG 165  
Db 49 LSGHSAVYTKFNPAAGTLIAGSGSHDREIFLWRVHGDKNFVWLGKHNAILDLHTSDG 108  
QY 166 KRLASGGKAGSIITWDPETGQOKRPLSGHKHINCIAWEYHRDPECRKLASAGGDC 225  
Db 109 SQIVSAPDKTVRAWDVETGKQI-KKMAEHSFVNSCC--PTRRGPPL--IISGDDGTA 163  
QY 226 RIWDVKLQCCIMNIAGHTNAVTAVRWG-GAGLIYTSKDRTVKWMRAADGILCTFSGHA 284  
Db 164 KLWDMRQGAITQTFDPKYO-ITAVSFDAADKIFTGGVNDVKVWDLRKGEATWTLEGHQ 222  
QY 285 HWNNIALSTD--YVLRGTGPHFPVKDRSKSHLSISTEELQESALARYQAVCPDEVELYS 342  
Db 223 DTITGNSLSPDGSLITNGM----- 242  
QY 343 CSDDNTLYLMRNN-----QNKCVERTGHQNVVNDVYSPDVKLIASAFKSVRLWRASD 398

Db 243 ---DNKLCVMDMRPYAFQNRCKVIFEGHQH-----NEKN----- 274  
QY 399 GOYMATFRGHYQAVTVVAWSADSRLLIVSGSKDSTLKVMSVQTKKLAQELPGHAEVFGVD 458  
Db 275 -----LLKGSWSPDGTKYTAGSSDBWVHIWDTTSRRTYKLPGLHTGVSNECV 321  
QY 459 WAPDGSRVASGGKDKVIKL 477  
Db 322 FHTEPLIGSCSSDKNIYL 340

Search completed: June 2, 2003, 07:03:21  
Job time : 37 secs





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 2, 2003, 06:59:40 ; Search time 264 Seconds  
(without alignments)  
184.043 Million cell updates/sec

Title: US-09-830-980-1  
Perfect score: 2574  
Sequence: 1 MGETTEQATPHIQARLV.....PDGSRVASGGKVKIKLWAY 480

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 383519 seqs, 101223694 residues  
Total number of hits satisfying chosen parameters: 383519

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published\_Applications\_AA:\*  
1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US05\_NEW\_PUB pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1521	59.1	485	9	US-10-132-744A-6
2	1452	56.4	484	9	US-10-132-744A-2
3	819	31.8	261	9	US-10-132-744A-4
4	581	22.6	1356	9	US-10-077-111-10
5	409.5	15.9	742	9	US-10-077-111-11
6	348	13.5	521	10	US-09-764-853-449
7	346	13.4	678	10	US-09-801-368-314
8	344	13.4	159	9	US-09-774-639-247
9	344	13.4	159	9	US-09-969-730-345
10	323.5	12.6	343	9	US-10-119-932-5
11	320	12.4	423	10	US-09-729-674-160
12	313.5	12.2	569	9	US-10-038-010-8
13	313.5	12.2	569	12	US-10-042-417-2
14	312	12.1	1194	10	US-09-876-667-2
15	312	12.1	1205	10	US-09-876-667-16
16	310.5	12.1	640	9	US-10-060-019-30
17	308	12.0	375	9	US-10-119-932-1
18	302	11.7	540	9	US-09-213-888-7
19	302	11.7	540	9	US-09-213-888-10

20	302	11.7	540	9	US-09-328-877A-7	Sequence 7, Appli
21	302	11.7	540	9	US-09-328-877A-10	Sequence 10, Appli
22	302	11.7	545	9	US-09-213-888-6	Sequence 6, Appli
23	302	11.7	545	9	US-09-328-877A-6	Sequence 6, Appli
24	302	11.7	553	9	US-09-213-888-5	Sequence 5, Appli
25	302	11.7	553	9	US-09-328-877A-5	Sequence 5, Appli
26	302	11.7	559	9	US-09-213-888-9	Sequence 9, Appli
27	302	11.7	559	9	US-09-328-877A-9	Sequence 9, Appli
28	302	11.7	589	9	US-09-213-888-8	Sequence 8, Appli
29	302	11.7	589	9	US-09-328-877A-8	Sequence 8, Appli
30	302	11.7	592	9	US-09-213-888-4	Sequence 4, Appli
31	302	11.7	592	9	US-09-328-877A-4	Sequence 4, Appli
32	302	11.7	626	9	US-09-213-888-21	Sequence 21, Appli
33	302	11.7	626	9	US-09-328-877A-21	Sequence 21, Appli
34	302	11.7	627	9	US-09-213-888-3	Sequence 3, Appli
35	302	11.7	627	9	US-09-328-877A-3	Sequence 3, Appli
36	302	11.7	666	9	US-09-213-888-27	Sequence 27, Appli
37	302	11.7	666	9	US-09-328-877A-27	Sequence 27, Appli
38	302	11.7	669	9	US-09-213-888-25	Sequence 25, Appli
39	302	11.7	669	9	US-09-328-877A-25	Sequence 25, Appli
40	299	11.6	613	9	US-10-108-605-181	Sequence 181, App
41	294.5	11.4	713	10	US-09-801-368-408	Sequence 408, App
42	292.5	11.4	402	10	US-09-925-300-1320	Sequence 1320, Ap
43	292.5	11.4	404	9	US-09-943-689A-2	Sequence 2, Appli
44	289.5	11.2	400	9	US-09-943-689A-4	Sequence 4, Appli
45	288.5	11.2	404	10	US-09-908-805B-21	Sequence 21, Appli

ALIGNMENTS

RESULT 1  
US-10-132-744A-6  
; Sequence 6, Application US/10132744A  
; Publication No. US20030027261A1  
; GENERAL INFORMATION:  
; APPLICANT: Utku, Nalan  
; TITLE OF INVENTION: No. US20030027261A1 genes Tzap7/A, Tzap7/B and Tzap7 invol  
; TITLE OF INVENTION: activation and uses thereof  
; FILE REFERENCE: Utku-4 CON  
; CURRENT APPLICATION NUMBER: US/10/132,744A  
; PRIOR FILING DATE: 2002-04-24  
; PRIOR APPLICATION NUMBER: PCT/EP00/10670  
; PRIOR FILING DATE: 2000-10-30  
; PRIOR APPLICATION NUMBER: 60/185,016  
; PRIOR FILING DATE: 2000-02-23  
; PRIOR APPLICATION NUMBER: 60/162,675  
; PRIOR FILING DATE: 1999-11-01  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 485  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; NAME/KEY: miscellaneous feature  
; LOCATION: 379  
; OTHER INFORMATION: variable amino acid  
US-10-132-744A-6

Query Match 59.1%; Score 1521; DB 9; Length 485;  
Best Local Similarity 58.6%; Pred. No. 2.7e-123;  
Matches 279; Conservative 73; Mismatches 118; Indels 6; Gaps 3;  
QY 9 EATPHITQARLVYTGEEA---GPPIDLPAGITITQGLICNALLKNEATPVLFVCGED 64  
DB 8 EAVARDVQELLVQFDGGQILGSPFDPVDITPRLQVLVCNALLAQEDPLAFVHDA 67  
QY 65 ETKKSLDTLDLASVDYENVIDIVYQPAVFKVRVTRCTSSMPGHAEAVSLNFSPOGA 124  
DB 68 EIVSSLGKLTLESQAVETEKVLDIYQPAIFRVAIVTRCTSSLEHSEAVISVAFSPGK 127  
QY 125 HLAGSGDITVRLWDLNLTETPHFTCTGKQKQVLCVSNAPDKRLASCKAGSIIDWDPET 184

Db 128 YLASGGDTTVRFWDLSTETPHFTCKGHRHWLVLSISWPDGKKLASGCKNGQIILWDPST 187  
Qy 185 GQOKRPLSGHKKHINCLAWEPYHRDPECRKLASASGDGCRIDWVKLGQCLMNIAGHTN 244  
Db 188 GKQVRLTLAGHKKWITGLSWEPLHANPECRVYVASSKDGSRVLDWTAGRCERILTGHTQ 247  
Qy 245 AVTAVRWGAGLIYTSKORTVKMWRADGILCFTFSGHAHWNNIALSTDYVLRTPGFH 304  
Db 248 SVTCLRWGGDLGLYSASQDRTIKVWRAHDGVLCTLGQGHGHWNTMALSTDYALRTGAFE 307  
Qy 305 PVK-DRSKSHLSLSTEELOESALKRYQAVCPDEVESLVCSDNTLYLWRNNQ-KCVER 362  
Db 308 PAEASVNPQDLOGSLQELKERALSRYNLVRGOGPERLVSGSDDTFLWSPAEDKKPLTR 367  
Qy 363 MTGHONVNDVKYSPDKLIASASDFKSVRLWRASDGQYMATFRGHVQAVTVVWASDSR 422  
Db 368 MTGHQALINQVYSPDSRIVASASDFKSIKLDWGRGTGKYLASLRGHVAVYQIAWSADR 427  
Qy 423 LIVSGSKDSTLKWMSVOTKKLAQELPGHADEVFGVWAPDGSRVASGGKDKVILW 478  
Db 428 LTVSGSDSTLKVMDYKAQKLAIDLPGHADEVAVDWSPDGQVRVSGGKDKCLRIW 483

## RESULT 2

US-10-132-744A-2  
; Sequence 2, Application US/10132744A  
; Publication No. US20030027261A1  
; GENERAL INFORMATION:  
; APPLICANT: Utku, Nalan  
; TITLE OF INVENTION: NO. US20030027261A1el genes Tzap7/A, Tzap7/B and Tzap7 Involved  
; FILE REFERENCE: Utku-4 CON  
; CURRENT APPLICATION NUMBER: US/10/132,744A  
; PRIOR FILING DATE: 2002-04-24  
; PRIOR APPLICATION NUMBER: PCT/EP00/10670  
; PRIOR FILING DATE: 2000-10-30  
; PRIOR FILING DATE: 2000-02-25  
; PRIOR APPLICATION NUMBER: 60/185,016  
; PRIOR FILING DATE: 2000-02-25  
; PRIOR APPLICATION NUMBER: 60/162,675  
; PRIOR FILING DATE: 1999-11-01  
; NUMBER OF SEQ ID NOS: 2.1  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 2  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-132-744A-2

Query Match 56.4%; Score 1452; DB 9; Length 484;  
Best Local Similarity 57.6%; Pred. No. 2.6e-117; Mismatches 130; Indels 6; Gaps 3;  
Matches 274; Conservative 66; Mismatches 130; Indels 6; Gaps 3;

Qy 9 EATPHIOARLVYTGEEA---GPPIDLPAGITTTQQLGLICNALKNKEEATPYLFFVYGD 64  
Db 7 EAVADVORLLVQPDGEGQLLSPDFVVDITPDRLQLVCNALLAQEDPLSLAFFVHDA 66  
Qy 65 EIKKSLEDTDLASVDTEENVIDIVYQPAVFKVPEVTRCTSSMPGHAFAVYSLNFSPDGA 124  
Db 67 EIVSLSGKTLSEQAVETEKVLDIILPTQAVFKVAVTRCTISSLEGTEAVISVAFSTGK 126  
Qy 125 HLASGGDTTVRLWDLNTEPHFTCTGHQWVLCVSWAPDGRKLAGSKAGSIILWDPET 184  
Db 127 YLASGGDTTVRFWDLSTETPHFTCKGHRHWLVLSIAWSPDGKKLASGCKNGSQIFWDPST 186  
Qy 185 GQOKRPLSGHKKHINCLAWEPYHRDPECRKLASASGDGCRIDWVKLGQCLMNIAGHTN 244  
Db 187 GKQIKPLTGHKKWITGLSWEPLHANPECRVYVASSKDGSRVLDWTAGRCERILTGHTQ 246  
Qy 245 AVTAVRWGAGLIYTSKORTVKMWRADGILCFTFSGHAHWNNIALSTDYVLRTPGFH 304  
Db 247 SVTCLRWGGDLGLYSASQDRTIKVWRAHDGVLCTLGQGHGHWNTMALSTDYALRTGAFE 306

Qy 305 PVK-DRSKSHLSLSTEELOESALKRYQAVCPDEVESLVCSDNTLYLWRNNQ-KCVER 362  
Db 307 PAEASVNPQDLOGSLQELKERALSRYNLVRGOGPERLVSGSDDTFLWSPAEDKKPLTR 366  
Qy 363 MTGHONVNDVKYSPDKLIASASDFKSVRLWRASDGQYMATFRGHVQAVTVVWASDSR 422  
Db 367 MTGHQALINQVYSPDSRIVASASDFKSIKLDWGRGTGKYLASLRGHVAVYQIAWSADR 426  
Qy 423 LIVSGSKDSTLKWMSVOTKKLAQELPGHADEVFGVWAPDGSRVASGGKDKVILW 478  
Db 427 LTVSGSDSTLKVMDYKAQKLAIDLPGHADEVAVDWSPDGQVRVSGGKDKCLRIW 482

## RESULT 3

US-10-132-744A-4  
; Sequence 4, Application US/10132744A  
; Publication No. US20030027261A1  
; GENERAL INFORMATION:  
; APPLICANT: Utku, Nalan  
; TITLE OF INVENTION: NO. US20030027261A1el genes Tzap7/A, Tzap7/B and Tzap7 Invo  
; FILE REFERENCE: Utku-4 CON  
; CURRENT APPLICATION NUMBER: US/10/132,744A  
; PRIOR FILING DATE: 2002-04-24  
; PRIOR APPLICATION NUMBER: PCT/EP00/10670  
; PRIOR FILING DATE: 2000-10-30  
; PRIOR APPLICATION NUMBER: 60/185,016  
; PRIOR FILING DATE: 2000-02-25  
; PRIOR APPLICATION NUMBER: 60/162,675  
; PRIOR FILING DATE: 1999-11-01  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 261  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-132-744A-4

Query Match 31.8%; Score 819; DB 9; Length 261;  
Best Local Similarity 58.8%; Pred. No. 6.3e-63;  
Matches 150; Conservative 41; Mismatches 62; Indels 2; Gaps 2;

Qy 226 RIWDVKLGQCLMNIAGHTNAVTVRWGAGLIYTSKDRFTVKMWRADGILCFTFSGHAH 285  
Db 5 RIWDTTAGRCERILTGHTQSVTCLRWGGDLGLYSASQDRTIKVWRAHDGVLCTLGQGHG 64  
Qy 286 WYNNIALSTDYVLRTPGFHPPVK-DRSKSHLSLSTEELOESALKRYQAVCPDEVESLVCSS 344  
Db 65 WYNTMALSTDYALRTGAFEPAEASVNPQDLOGSLQELKERALSRYNLVRGOGPERLVSGS 124  
Qy 345 DDNTLYLWRNNQ-KCVERMTGHONVNDVKYSPDKLIASASDFKSVRLWRASDGQYMA 403  
Db 125 DDFTFLWSPAEDKKPLTRMTGHQALINQVLFSPDSRIVASASDFKSIKLDWGRGTGKYL 184  
Qy 404 TFRGHVQAVTVVWASDSRLYVSGSKDSTLKWMSVOTKKLAQELPGHADEVFGVWAPDG 463  
Db 185 SLRGHVAAYQIAWSADSRLLYVSGSDSTLKWMDYKAQKLAIDLPGHADEVAVDWSPDG 244  
Qy 464 SRVASGGKDKVILW 478  
Db 245 QRVASGGKDKCLRIW 259

## RESULT 4

US-10-077-111-10  
; Sequence 10, Application US/10077111  
; Publication No. US20020187492A1  
; GENERAL INFORMATION:  
; APPLICANT: Todderud, C. Gordon  
; APPLICANT: Finger, Joshua N.  
; APPLICANT: Rillema, Jill  
; TITLE OF INVENTION: TBA  
; FILE REFERENCE: 3053-4114US2



Db 119 ASASDRRTVRLWIPD---KROK----- 137  
Qy 229 DVKLQCLMNIAGHTNAVTVRMGGAG-LIYTSSKDRIVKWRADGILCFTFSGHAWV 287  
Db 138 -----FSEFKAHTAPVRSYDFADQGLFATASEDKSIKWSMYRQFLYSIRYHTWV 190  
Qy 288 NNIALSTDYVLRTPGPHVPKDRSKSHLSLSTELQESALKRYQAVCPDEVESLVSCSDN 347  
Db 191 -----RCAKFSF-----DGRLLVSCSEDK 209  
Qy 348 TLYLWRNNQKVERMTGHONVNDVKYSPDVKLTASAFDKSVRLWRASDGOYMATFRG 407  
Db 210 TIKTWDTTNKQCVNFSVSGFANFYDFNPSGTCFCLASAGSDQTVKWDVVRNKLQHYQV 269  
Qy 408 HQAVYTVAWSADSRLLVSGSKOSTLKVMSVQKKLAQELFGHADEVFGVDNAPGDSRVA 467  
Db 270 HSGGVNCSIFSPHSGNYLITASSDGTLLKILDLEGRLLYTLQGTGPVFTVSFKGGELFA 329  
Qy 468 SGGKDKVILW 478  
Db 330 SGGADTOVLLW 340

## RESULT 7

US-09-801-368-314  
; Sequence 314, Application US/09801368  
; Patent No. US20020128250A1  
; GENERAL INFORMATION:  
; APPLICANT: Busby, Robert  
; APPLICANT: Cali, Brian  
; APPLICANT: Hecht, Peter  
; APPLICANT: Holtzman, Doug  
; APPLICANT: Madden, Kevin  
; APPLICANT: Maxon, Mary  
; APPLICANT: Milne, Todd  
; APPLICANT: No. US20020128250A1man, Thea  
; APPLICANT: Rover, John  
; APPLICANT: Salama, Sofie  
; APPLICANT: Sherman, Amir  
; APPLICANT: Silva, Jeff  
; APPLICANT: Summers, Eric  
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi  
; FILE REFERENCE: 109272.147  
; CURRENT APPLICATION NUMBER: US/09/801.368  
; PRIOR FILING DATE: 2001-03-07  
; PRIOR FILING DATE: US 09/487,558  
; PRIOR FILING DATE: 2000-01-19  
; PRIOR FILING DATE: US 60/160,587  
; NUMBER OF SEQ ID NOS: 440  
; SOFTWARE: Patent in version 3.0  
; SEQ ID NO 314  
; LENGTH: 678  
; TYPE: PRT  
; ORGANISM: Aspergillus nidulans  
US-09-801-368-314

Query Match 13.4%; Score 346; DB 10; Length 678;  
Best Local Similarity 24.2%; Pred. No. 2e-21;  
Matches 95; Conservative 60; Mismatches 118; Indels 120; Gaps 11;  
Qy 148 TCTGKQWLVCSWAPDGKRLASGCKAGSIWIWDPTGQKGRPLSGHKKHINLAWEPY 207  
Db 344 TFKGHTNGVMCLQF--EDNILATGSDYTTIKIWDTEGEL-RTLRGHESGIRCLQFDD- 399  
Qy 208 HRDEPCKLASAGSDGDCRWDVKLGOC LMNIAGHNNAVTVRMGGAGLIVTSSKDRVTYK 267  
Db 400 -----TKLISGMDRTIKVNNWRTGECISTYTGHHGGVIGLHF-DASILASGVDKIVK 452  
Qy 268 MRAADGILCRIFS--GHAHWNNIALSTDYVLRTPGPHVPKDRSKSHLSLSTELQESA 325  
Db 453 IWNFECK---STFSLRGHTDWNVAVRVDT----- 478

Qy 326 LKRYQAVCPDEVESLVSCSDNTLYLWRNNQKVERMTGHONVNDV----- 373  
Db 479 -----SSRTVFSASDDCTVRLWDLDTKTCIRTFHGHVGOVQVPLPRFEFEH 528  
Qy 374 -----KYSPOVKLIASASF 387  
Db 529 DAECENUDLSTTSGDANPPFSIQASMGLEPNAAYSQSSAFGTSFSDNGRAAPRYMTSA-L 587  
Qy 388 DKSVRLWRASDGOYMATFRGHVQAVTVAWSADSRLLVSGSKOSTLKVMSVQKKLAQEL 447  
Db 588 DSTIRLWETTTGCLRTFFGHLEGVW--ALGADYLRIVSGAEDRMIKIWDPRTKCERTE 645  
Qy 448 PGHADEVFGVDNAPGDSRVSAGGKDKVILW 480  
Db 646 TGHSGPVTICIGL--DSRFATGSECEVRMTSF 676

## RESULT 8

US-09-774-639-247  
; Sequence 247, Application US/09774639  
; Publication No. US20030003555A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: 90 Human Secreted Proteins  
; FILE REFERENCE: P2013P1  
; CURRENT APPLICATION NUMBER: US/09/774,639  
; PRIOR FILING DATE: 2001-07-09  
; PRIOR FILING DATE: EARLIER APPLICATION NUMBER: 09/244,112  
; NUMBER OF SEQ ID NOS: 371  
; SOFTWARE: Patent in Ver. 2.0  
; SEQ ID NO 247  
; LENGTH: 159  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (63)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (137)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-774-639-247

Query Match 13.4%; Score 344; DB 9; Length 159;  
Best Local Similarity 53.3%; Pred. No. 3.9e-22;  
Matches 72; Conservative 20; Mismatches 41; Indels 2; Gaps 2;  
Qy 296 YVLRTPGPHVPK-DRSKSHLSLSTELQESALKRYQAVCPDEVESLVSCSDNTLYLWRN 354  
Db 1 YALRTGAPEPAEASVNPDDLQGSLOELKERALSRYNLVRGQGPRLVSGSDDTFLWSP 60  
Qy 355 NO-NKCVERTGHONVNDVKYSPDVKLTASAFDKSVRLWRASDGOYMATFRGHVQAVY 413  
Db 61 AEXKKFLTRMTGHQALLNQVLFSPDSRIVASASAFDKSIKLDGRTGKYLASIRGHVAAVY 120  
Qy 414 TVAWSADSRLLVSGS 428  
Db 121 QIAWSADSRLLVSGS 135

## RESULT 9

US-09-969-730-345  
; Sequence 345, Application US/09969730  
; Publication No. US2003005443A1  
; GENERAL INFORMATION:  
; APPLICANT: Ruben et al.  
; TITLE OF INVENTION: 90 Human Secreted Proteins  
; FILE REFERENCE: P2013P2  
; CURRENT APPLICATION NUMBER: US/09/969,730  
; PRIOR FILING DATE: 2001-10-04  
; PRIOR FILING DATE: 09/774,639  
; PRIOR FILING DATE: 2001-02-01





Db 484 VSGAYDGKIKVWDL-VAALDPRAPAGTLCUPLTVEHSGRVFRLQF--DEFQIVSSSHDDT 540  
QY 475 IKLWAY 480  
Db 541 ILIWD 546

RESULT 13  
US-10-042-417-2  
; Sequence 2, Application US/10042417  
; Patent No. US20020123082A1  
; GENERAL INFORMATION:  
; APPLICANT: Pagano, M.  
; TITLE OF INVENTION: METHODS TO IDENTIFY COMPOUNDS USEFUL FOR THE TREATMENT OF  
; FILE REFERENCE: 5914-090-999  
; CURRENT FILING DATE: 2002-01-07  
; PRIOR FILING DATE: 2001-01-05  
; NUMBER OF SEQ ID NOS: 89  
; SOFTWARE: Patent Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 569  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-042-417-2

Query Match 12.2%; Score 313.5; DB 12; Length 569;  
Best Local Similarity 24.6%; Pred. No. 1e-18;  
Matches 90; Conservative 59; Mismatches 132; Indels 85; Gaps 12;

QY 139 DLNTEPHFTCTGH-----KQWLCVSWAPDGRKLASGCKAGSIILWDPETGOO 187  
Db 242 DIETIESNRCGRHSQRHCRSETSKGVCLQY--DDQKIVSGLRDNTIKINDKTLFC 299  
QY 188 KGRPLSGHKKHINCLAWPEYHROPECKKLASAGSDGCRINDVKLGQCLMNIAGHTNATV 247  
Db 300 K-RILTHTGSLVCLQYDE-----RVITGSSDSTVRWDVNTGEMLNTLIHCEAVL 351  
QY 248 AVRWGAGLIYTSKORTVKMRAA---DGLICRTFSGHAHWNNIALSTDYVLRTPGFH 304  
Db 352 HLRFNN-GMVTCSKORSIAVWDMASPTDITLRLVLVGHRAAVNVVDFDDKI----- 403  
QY 305 PVKDRSKSHLSLSTELQESALKRYQAVCPDEVESLVSCSDNTLYLWRNNQKVCVERMT 364  
Db 404 -----VSASGDRTIKVNWTSTCEFEVRLN 427  
QY 365 GHQNVVNDVKYSPDVKLIASAFKSVRLMARSDGOYMATFRGHQVQAVYTVMSADSRLLI 424  
Db 428 GHRGIACLYRD--RLVVGSSDNTIRLWDIECGACLRVLGHEELVRCIRF--DNKRI 483  
QY 425 VSGSKDSTLKVSVQTKLAQLPG-----HAEVFGVDWAPDGSVAGSGKDKV 474  
Db 484 VSGAYDGKIKVWDL-VAALDPRAPAGTLCUPLTVEHSGRVFRLQF--DEFQIVSSSHDDT 540  
QY 475 IKLWAY 480  
Db 541 ILIWD 546

RESULT 14  
US-09-876-667-2  
; Sequence 2, Application US/09876667  
; Patent No. US20020107370A1  
; GENERAL INFORMATION:  
; APPLICANT: Henzel, William J.  
; TITLE OF INVENTION: APAF-1, AN ACTIVATOR OF C ASPASE-3  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Merchant, Gould, Smith, Edell, Welter & Schmidt  
; STREET: 3100 No. US20020107370A1west Center, 90 South Seventh St

CITY: Minneapolis  
STATE: MN  
COUNTRY: USA  
ZIP: 55402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/876,667  
FILING DATE: 07-Jun-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/435,115  
FILING DATE: <Unknown>  
APPLICATION NUMBER: 60/055,258  
FILING DATE: 07-AUG-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Kettelberger, Ph.D., Denise M  
REGISTRATION NUMBER: 33,924  
REFERENCE/DOCKET NUMBER: 11669.60SU1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 612-332-5300  
TELEFAX: 612-332-9081  
TELEX: <Unknown>  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1194 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FRAGMENT TYPE: internal  
SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
US-09-876-667-2

Query Match 12.18; Score 312; DB 10; Length 1194;  
Best Local Similarity 22.5%; Pred. No. 4e-18;  
Matches 130; Conservative 78; Mismatches 200; Indels 170; Gaps 20;  
QY 40 QLGLICNALLKNERATPYLFVGEDEIKKSLED-TLDLASVDTEENVIVQPAQVFKVR 98  
Db 554 QLGL-C-----EPETSEVYQAKLQAKQDEVDNGMLLEWINKNITNL-----SRLVVR 601  
QY 99 PVTRCTSSMPGHAENVSLNFSPDGAHLASGGDTTVRLMDLNTETPHETCTGCHKQWLC 158  
Db 602 P-----HTDAVYHACFSEDGQRIASCAGADKTLQVFKAEKLEIKAHEDVLC 651  
QY 159 VSWAPDGKELASGCKAGSIILWDPETGOQKGRPLSGHKKHINCLAWPEYHROPECKRLAS 218  
Db 652 CAPSTDDREIATCSVDKVKIWNMTGELY-HTVDEHSEQVNC-----HFTNSSHLL 705  
QY 219 ASGDGDC--RIWDVKLQCLMNTAGHTNATVAVRWG--GAGLIYTSKORTVKMRA----- 271  
Db 706 ATGSSDCFLLKWLMDLNQKCRNTMFGHTNSVNHCRFSPDDKLIASCSADGTLKMDATSAN 765  
QY 272 -----ADG-----ILCRTFS----- 281  
Db 766 ERKSNVKKOFFNLNLEDPQEDMEVIVKCCSWNSADGARINVAANKIFLWNTDSRSKVDACR 825  
QY 282 GHAWVNNIALSTD-----YVL- 298  
Db 826 GHLVWVHGVFMFSPDGSFLTSSDDQITRLWETKVKCKNSAVMLKQEVVDVVFQENEVWVLA 885  
QY 299 -----RGPFPKPKDRSKSHLSLSTE-----ELOESAL--KR 328  
Db 886 VDIRRLQLINGRTQGDYITEAQVSCCCLSPHLYQYAFGDENGATEILELNNKIRFQSR 945  
QY 329 YQ-----AVCPDEVESLVSCSDNTLYLWRNNQKVCVERMTGHTNVVNDVKYSPDKV 380  
Db 946 FOHKKTVWHIQTFADE-KTLISSDDAEIQVNNWQDKCI-FLRGHQTVDKFDRLKNSR 1003





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 2, 2003, 06:54:30 ; Search time 45 seconds  
(without alignments)  
1025.434 Million cell updates/sec

Title: US-09-830-980-1

Perfect score: 2574

Sequence: 1 MQETDEQATPHIQARLV.....PDGSRVASGKDKVILWAY 480

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 73: \*  
1: PIR1: \*  
2: PIR2: \*  
3: PIR3: \*  
4: PIR4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1318.5	51.2	473	2 T33805	hypothetical prote
2	1092	42.4	515	2 S19487	hypothetical prote
3	1047.5	40.7	502	2 T41148	tip-asp repeat con
4	599	23.3	1258	2 AI2155	WD-repeat protei
5	597	23.2	1683	2 AF2071	WD-40 repeat prote
6	590	22.9	1526	2 AC2239	WD-40 repeat prote
7	581	22.6	1356	2 T18521	beta transducin-li
8	538.5	20.9	1708	2 AE1866	WD-40 repeat prote
9	533	20.7	1227	2 AE1810	WD-40 repeat prote
10	532.5	20.7	1711	2 AD1842	WD-40 repeat prote
11	532	20.7	1747	2 AC1842	WD-40 repeat prote
12	497	19.3	934	2 AG1889	WD-40 repeat prote
13	496.5	19.3	1189	2 AI2493	WD-repeat protei
14	486	18.9	1551	2 AB2410	WD-repeat protei
15	476.5	18.5	1676	2 AH2195	hypothetical prote
16	469	18.2	1189	2 AH2154	hypothetical prote
17	451	17.5	677	2 AE1861	serine/threonine k
18	436	16.9	304	2 AG1837	WD-40 repeat prote
19	433	16.8	1693	2 S76086	beta transducin-li
20	425.5	16.5	1049	2 T42045	hypothetical prote
21	424.5	16.5	559	2 AB2202	WD-40 repeat regul
22	410.5	15.9	317	2 T46032	WD-repeat protei
23	400	15.5	589	2 AG2400	hypothetical prote
24	372	14.5	376	2 T19266	WD-40 repeat prote
25	371	14.4	786	2 AG2375	hypothetical WD-re
26	368.5	14.3	777	2 T41075	WD40-repeat protei
27	365.5	14.2	876	2 T51507	hypothetical prote
28	361.5	14.0	501	2 T27513	WD-40 repeat regul
29	359	13.9	614	2 S58306	

30	355	13.8	333	2 G85034	probable WD-repeat
31	354	13.8	409	2 S36113	l1S-1 protein - hu
32	354	13.8	410	2 S48052	platelet-activatin
33	347.5	13.5	395	2 T23317	hypothetical prote
34	344.5	13.4	265	2 AF1890	WD-repeat protei
35	327.5	12.7	357	2 AI2099	WD-40 repeat prote
36	323.5	12.6	343	2 C84870	probable splicing
37	322	12.5	342	2 AE2490	WD-repeat protei
38	318.5	12.4	586	2 T38992	WD-40 repeat regul
39	314.5	12.2	518	2 B48088	beta-transducin re
40	314.5	12.2	598	2 AE2415	WD-repeat protei
41	312.5	12.1	323	2 T02617	hypothetical prote
42	312.5	12.1	605	2 T38932	probable sulfur me
43	312.5	12.1	701	2 T16607	hypothetical prote
44	312	12.1	1194	2 T03818	apoptotic proteina
45	310.5	12.1	640	2 S49932	MEF30 protein - ye

#### ALIGNMENTS

##### RESULT 1

T33805

hypothetical protein W07E6.2 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 26-May-2000

C:Accession: T33805

R:Latreille, P.; Wamsley, P.

submitted to the EMBL Data Library, November 1998

A:Description: The sequence of C. elegans cosmid W07E6.

A:Reference number: Z21414

A:Accession: T33805

A:Status: preliminary; translated from GB/EMBL/DD8J

A:Molecule type: DNA

A:Residues: 1-473 <LAT>

A:Cross-references: EMBL:AF106576; PIDN:AACT8176.1; GSPDB:GN00020; CESP:W07E6.2

A:Experimental source: strain Bristol N2; clone W07E6

C:Genetics:

A:Gene: CESP:W07E6.2

A:Map position: 2

A:Introns: 46/2; 77/3; 103/1; 195/2; 256/3; 311/3; 399/2

C:Superfamily: unassigned WD repeat proteins; WD repeat homology

Query Match	51.2%	Score	1318.5	DB	2	Length	473
Best Local Similarity	53.8%	Pred. No.	9e-95				
Matches	250	Conservative	72	Mismatches	122	Indels	21
Gaps	7						
QY	24	EEAGPPTDLPAGITTTQQLGLICNALL-KNEEATPYLFFVGE----	DEIKKSLDFTDLA	77			
DB	18	ELGSGGLVPVDISTNELIQLNQLGSSDDPVPISFFTTGAEIVDSIRKSLER----	72				
QY	78	SVDTENVVDIVYQPAVFKVPVTRCTSSMGHAAVYSLMFSPDGAHLASGSGDTTVRL	137				
DB	73	-IDFETTLKLVQPAVFRVPTKCSASIPGHGPVISAQFSDRGSLASGSGDTTVRL	131				
QY	138	WDLNTEPHTCTGKHQWLCVSWADGKRLASGCGKAGSIIDPPTGQOKRPLSGHKK	197				
DB	132	WDIELEPLTKSKHKNVLCIAWSPDATKIASCKNGEICINAKTGGQIGTKLRKQ	191				
QY	198	HINCLAWEPYHRDPCRKLASGSGDCRIWDVKLGQCCLMNIAGHNNAVAVWVGAGLI	257				
DB	192	WITSLAWQPMHDKPTCRLLASCGKGNIFITWTVGTGTVVYRCLSGHTASVTCURWEGGLI	251				
QY	258	YTSSKORTVMKRAADGILCTFSGHAAVNNIALSTDYVLRTPGPHFKDRSKSHLSLS	317				
DB	252	YSGQDRTVMKMRADGGVMCRNMTGHAHNTLALNTDYLARTSCFEPKRCIKPD----	308				
QY	318	TEELOESALKRYQAVCPDEV---ESLVSGSDNTLYLRNNQNK-CVERMTGHQVNDV	373				
DB	309	VEECQKVAQTRYEAL--EIAGGERLVSGSDDTFLFMNPKETKQSIINRTMTGMQVNVQV	366				
QY	374	KYSPDVKLTASDFKVSFWRASDCQYATFGRVAVYTVANSGLVYSGSKDSTL	433				

Db 367 VSPDTRYIASASFDKSVKLWCGRTGKYLASLRCHVGPVYQVANSADSRLLVSGSADSTL 426  
QY 434 KWSVQVKKLAOELPGHAEVGVWDWADGSRVASGGKDKVILKW 478  
Db 427 KVFELTKSLYDLPGHGEVFTVDMSPGKTKVVSAGKDKVILKW 471

## RESULT 2

S19487  
hypothetical protein YCR072c - yeast (Saccharomyces cerevisiae)  
C:Species: Saccharomyces cerevisiae  
C:Date: 31-Mar-1992 #sequence\_revision 04-Dec-1992 #text\_change 19-Apr-2002  
C:Accession: S19487; S26657  
R:Ballesta, J.P.G.; Franco, L.; Hoenicka, J.; Jimenez, A.; Remacha, M.; Sanz, E.  
submitted to the Protein Sequence Database, March 1992  
A:Reference number: S19486  
A:Accession: S19487  
A:Molecule type: DNA  
A:Residues: 1-484, 'I', 486, 'TKL', 490, 492-515 <BAL1>  
A:Cross-references: EMBL:X59720; MIPS:YCR072c  
A:Note: this sequence has been revised in reference S26657  
R:Ballesta, J.P.G.; Franco, L.; Hoenicka, J.; Jimenez, A.; Remacha, M.; Sanz, E.  
submitted to the Protein Sequence Database, October 1992  
A:Reference number: S26587  
A:Accession: S26657  
A:Molecule type: DNA  
A:Residues: 481-503 <BAL2>  
A:Cross-references: EMBL:X59720; MIPS:YCR072c  
C:Genetics:  
A:Map position: 3R  
A:Superfamily: unassigned WD repeat proteins; WD repeat homology  
F:182-215/Domain: WD repeat homology <WD2>  
F:225-265/Domain: WD repeat homology <WD3>  
F:274-306/Domain: WD repeat homology <WD4>  
F:398-431/Domain: WD repeat homology <WD5>  
F:440-473/Domain: WD repeat homology <WD6>

Query Match 42.4%; Score 1092; DB 2; Length 515;  
Best Local Similarity 43.3%; Pred. No. 4.3e-77;  
Matches 214; Conservative 95; Mismatches 147; Indels 38; Gaps 11;

QY 14 TIQARLVYTGEGAPPIDLPAGITTOQLGLICNALL-KNEEATPYLFV-----GEDI 66  
Db 33 SIKFOALDTGDNVGGALRVPGAISEKQLELLNQLNGTSDDPPVYFTSCIIQKKASDV 92  
QY 67 KKSLEDITDLAS-----VDENVIDIVYQVAVKVPYTRCTSSMPGHAEAV 115  
Db 93 K-----TIDITDNLISLKPQYNSTEDQITLTPRAVKVPYTRSSAIGRSTIL 147  
QY 116 SLNFSF-DGAHLASGGDTTVRLDLNTEPHFTCTGKQWVLCVSNAPDGKRLASGCKA 174  
Db 148 CSAFAPHTSSRMVTGAGDNTARINDCDTQTPMHTLKGHNWVLCVSNAPDGEVIATGSD 207  
QY 175 GSIIWDPTGQGRPLSKHKKHINCIAWEPYH--RDPECKRLASASGGDCRIWDVKL 232  
Db 208 NTRILWDPKGQCIQDALRGHSHKWTLSWEPILHVKPKSKPLASSKDGTIKINDTVS 267  
QY 233 GQCLANIAGHFNVAIVRWGGAGLIYTSKDRTVKMWRA-ADGILCRFTSGHAHWNNIA 291  
Db 268 RVCQYTMSTGHTNSVCVKGWGGOLLYSGSHDRVTVVINSOGRCINILKSHAHWNHLS 327  
QY 292 LDCDVLVLTGPF-HPVKDRSKSHLSLSTPELOESALKRYQAVCP---DEVESLVSGSDN 347  
Db 328 LSTDVALIGAFDHTGKPS-----TPEAQKALENYEKICKNGNSEMMVATASDDY 381  
QY 348 TLYLNRN-NQNKVERMTGHQVNDVYSPDKVLASFDKSVRLWRASDQYWAIFR 406  
Db 382 TMLFNWPLKSTKPIARTGHQVNVHAFSPDGRYIVSASFNSIKLWDGRGCKFISTR 441  
QY 407 GHVQVYTVAWSADSRLLVSGSKDSTLKWSVOTKKLAQELPGHAEVGVWDWADGSRV 466  
Db 442 GHIASVYQVANSDCRLLVSGSKDTTLKWDVTRTKLSVDLEPGHKDEYITVDWSVDGKR 501

QY 467 ASGGKDKVILKW 480  
Db 502 CSGGKDKVRLWTH 515

## RESULT 3

T41148  
trp-asp repeat containing protein - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 26-May-2000  
C:Accession: T41148  
R:Hilbert, H.; Duesterhoeft, A.; Wood, V.; Rajandream, M.A.; Barrell, B.G.  
submitted to the EMBL Data Library, October 1998  
A:Reference number: Z21973  
A:Accession: T41148  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-502 <HIL>  
A:Cross-references: EMBL:AL031907; PIDN:CAA21419.1; GSPDB:GN00068; SPDB:SPCC18.05c  
A:Experimental source: strain 972h-; cosmid c18  
C:Genetics:  
A:Gene: SPDB:SPCC18.05c  
A:Map position: 3  
A:Introns: 68/3; 200/3; 299/3  
C:Superfamily: unassigned WD repeat proteins; WD repeat homology

Query Match 40.7%; Score 1047.5; DB 2; Length 502;  
Best Local Similarity 45.3%; Pred. No. 1.2e-73;  
Matches 208; Conservative 76; Mismatches 160; Indels 15; Gaps 6;

QY 32 LPAGITTOQLGLICNALLKN-BEATPYLFFVGEDEIKKSLDTDLA-----SVDTENV 85  
Db 49 VPGNSSVRQLEALLNQLNLSDDPPVNFALHHEDEFTIQTDLNLYTSVFHNLKMTEDHL 108  
QY 86 DIVYQVAVKVPYTRCTSSMPGHAEAVVSLNFSF-DGAHLASGGDTTVRLWDLNTE 144  
Db 109 TLLYTPQAVFRVRAVTRCTASMGHDGTIIISAQFSPSTSRLLVSGDFTARLWDCDTQT 168  
QY 145 PHFTCTGHKQWVLCVSNAPDGKRLASGCKAGSIINDPDTGQGRPLSGHKKHINCLAW 204  
Db 169 PIATMKGTNNVSCVAVAPDASIIATGSMNTIRFMDPKKSPIGDALRRHTKPIALCW 228  
QY 205 EYHDPCECRK--LASASGGDCRIWDVKLGQCLIMNTAGTNAVTAVRWGGAGLIYTSK 262  
Db 229 QPLHAPDSGPIYLLASGSKDNTVRVNWVLTLLFTLSGHTAPITCVWGQGNWYSSY 288  
QY 263 DRTVWMTAADGILCRFTSGHAHWNNIALSTOYLRTGTPHPVDRSKSHLSLSTEEQL 322  
Db 289 DKTIRWDKDKGCLHILKGHAARVNHLSLSTELHLSGAYDHTDFKSP-----SDER 343  
QY 323 ESALKRYQAVCPDEVESLVSCSDNTLYLM-RNNQNKVERMTGHQVNDVYSPDKVL 381  
Db 344 RKAKERYEACLKSGERLVSASDDQLDLWDPOKSTKPTIKMHGQKVVNHSFSPDGR 403  
QY 382 IASAFDKSVRLWRASDQYMATFRGHVQAVYTVAWSADSRLLVSGSKDSTLKWSVQTK 441  
Db 404 IATASFDSVRLWDGKTGFLATLRGHVAAVYQAVCANSTDRLLVSSQDTTLKWDVRSK 463  
QY 442 KLAOLPGHADVGVWDWADGSRVASGGKDKVILKW 480  
Db 464 KMKFDLPCHGEDQVFAVWDSPDQGVASGGADKAVRWIS 502

## RESULT 4

AI2155  
WD-repeat protein [imported] - Nostoc sp. (strain PCC 7120)

C:Species: Nostoc sp.  
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 30-Jun-2002  
C:Accession: AI2155  
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Nakazaki, N.; Shimpou, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.;

DNA Res. 8, 205-213, 2001  
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena sp. strain PCC 7120  
 A:Reference number: AB1807; MUID:21595285; PMID:11759840  
 A:Accession: AF2071  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1258 <KUR>  
 A:Cross-references: GB:BA000019; PIDN:BA74499.1; PID:gl7131893; GSPDB:GN00179  
 A:Experimental source: strain PCC 7120  
 C:Genetics:  
 A:Gene: alr2800

Query Match 23.3%; Score 599; DB 2; Length 1258;  
 Best Local Similarity 32.7%; Pred. No. 3.6e-38;  
 Matches 131; Conservative 66; Mismatches 150; Indels 54; Gaps 7;  
 QY 103 CTSSMPGHAEEAVVSLNFSPDGAHLASGSDTTVRLMDLNTETPHFTCTGCHKOW 155  
 DB 1061 EMOERNR---LEGHKDGVISISISRDQOTIASGSLDKTKLMSRDL-FRTLNGHEDA 1115  
 QY 156 VLYCVSWAPDCKKLASCAGSIIIDWPTGQGRPLSGHKKHINCLAWEPYHRDPECRK 215  
 DB 1116 VYVSFSPDQITIRSGGSDTKIKLQTSIDGTL-KYIIGHEQTNNV-----YFSPDGN 1169  
 QY 216 LASASGDCRWDVKLGQCLMNTAGHTNATVAVRWGAG-LIYTSKQRTVKWRAADG 274  
 DB 1170 LASASDHSIKLMDTTSGQLLTLTGHSAGVITVRFSPDQTIAGSEDKTVKLWHRDG 1229  
 QY 275 ILCTFSGSHAHVNNIALSTD-----YVLRTPGFPHVVKDRSKSHLSLSTELQESAL 295  
 DB 1230 KLLATLNGHODWVNSLSFSPDGKTLASASADTKIKLRIADGKLVKTLGHNDSDVMDVNF 1289  
 QY 296 -----YVLRTPGFPHVVKDRSKSHLSLSTELQESAL 327  
 DB 1290 SSDOKATASASRDNTIKLWNRHGLETFTHGSHGGVYAVNPLPDSNIIASASLDNTIRLW 1349  
 QY 328 RYQAVCPDEV-----YVLRTPGFPHVVKDRSKSHLSLSTELQESAL 337  
 DB 1350 QRPLISPLEVLGNSGYAVSFLHDGSIITAGADGNIQLWHSQDGSILKTLPGNKATYG 1409  
 QY 338 -----ESLVSCSDNTLYLRNNQKVERMTGHQVNVNDVYSPDKVLASASFPKS 390  
 DB 1410 ISFTPOGDILIASANADKTVKIVRVDRGKALKTLIGHDNEVKNVFPDGRKTLASASRDNT 1469  
 QY 391 VRLWRASDGOYMATFRGHVQAVTVVAMSADSLIYVSGKSTLKVWSVOTKKLAQELPGH 450  
 DB 1470 VKLWNSDGRKFKTLKGTDETFVFWVSFDPGKIIASASADTKIRLWDSFSGNLIKSLPAH 1529  
 QY 451 ADEVFVGVDAPDGSRVASGGRDKVILW 478  
 DB 1530 NDLYSVNFPDGSMLASTADKTVKLM 1557

RESULT 5  
 AF2071  
 WD-40 repeat protein [imported] - Nostoc sp. (strain PCC 7120)  
 C:Species: Nostoc sp.  
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
 C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 30-Jun-2002  
 C:Accession: AF2071  
 R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.  
 DNA Res. 8, 205-213, 2001  
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena sp. strain PCC 7120  
 A:Reference number: AB1807; MUID:21595285; PMID:11759840  
 A:Accession: AF2071  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1683 <KUR>  
 A:Cross-references: GB:BA000019; PIDN:BA73823.1; PID:gl7131215; GSPDB:GN00179  
 A:Experimental source: strain PCC 7120  
 C:Genetics:  
 A:Gene: al12124

Query Match 23.2%; Score 597; DB 2; Length 1683;  
 Best Local Similarity 26.2%; Pred. No. 7.6e-38;  
 Matches 149; Conservative 91; Mismatches 174; Indels 154; Gaps 9;  
 QY 36 ITFOGLIGNALLKNEATPYLFVGEDEIKSLIEDTLIASYDTENVIDIYQPAVF 95  
 DB 1019 LSHQQLAALASLKAQQNVHVI-----AVFNKLATVIT-----LQALF 1060

QY 96 KVRVTRCTSSMPGHAEEAVVSLNFSPDGAHLASGSDTTVRLMDLNTETPHFTCTGCHKOW 155  
 DB 1061 EMOERNR---LEGHKDGVISISISRDQOTIASGSLDKTKLMSRDL-FRTLNGHEDA 1115  
 QY 156 VLYCVSWAPDCKKLASCAGSIIIDWPTGQGRPLSGHKKHINCLAWEPYHRDPECRK 215  
 DB 1116 VYVSFSPDQITIRSGGSDTKIKLQTSIDGTL-KYIIGHEQTNNV-----YFSPDGN 1169  
 QY 216 LASASGDCRWDVKLGQCLMNTAGHTNATVAVRWGAG-LIYTSKQRTVKWRAADG 274  
 DB 1170 LASASDHSIKLMDTTSGQLLTLTGHSAGVITVRFSPDQTIAGSEDKTVKLWHRDG 1229  
 QY 275 ILCTFSGSHAHVNNIALSTD-----YVLRTPGFPHVVKDRSKSHLSLSTELQESAL 295  
 DB 1230 KLLATLNGHODWVNSLSFSPDGKTLASASADTKIKLRIADGKLVKTLGHNDSDVMDVNF 1289  
 QY 296 -----YVLRTPGFPHVVKDRSKSHLSLSTELQESAL 327  
 DB 1290 SSDOKATASASRDNTIKLWNRHGLETFTHGSHGGVYAVNPLPDSNIIASASLDNTIRLW 1349  
 QY 328 RYQAVCPDEV-----YVLRTPGFPHVVKDRSKSHLSLSTELQESAL 337  
 DB 1350 QRPLISPLEVLGNSGYAVSFLHDGSIITAGADGNIQLWHSQDGSILKTLPGNKATYG 1409  
 QY 338 -----ESLVSCSDNTLYLRNNQKVERMTGHQVNVNDVYSPDKVLASASFPKS 390  
 DB 1410 ISFTPOGDILIASANADKTVKIVRVDRGKALKTLIGHDNEVKNVFPDGRKTLASASRDNT 1469  
 QY 391 VRLWRASDGOYMATFRGHVQAVTVVAMSADSLIYVSGKSTLKVWSVOTKKLAQELPGH 450  
 DB 1470 VKLWNSDGRKFKTLKGTDETFVFWVSFDPGKIIASASADTKIRLWDSFSGNLIKSLPAH 1529  
 QY 451 ADEVFVGVDAPDGSRVASGGRDKVILW 478  
 DB 1530 NDLYSVNFPDGSMLASTADKTVKLM 1557

RESULT 6  
 AC2239  
 WD-40 repeat protein [imported] - Nostoc sp. (strain PCC 7120)  
 C:Species: Nostoc sp.  
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
 C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 30-Jun-2002  
 C:Accession: AC2239  
 R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.  
 DNA Res. 8, 205-213, 2001  
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena sp. strain PCC 7120  
 A:Reference number: AB1807; MUID:21595285; PMID:11759840  
 A:Accession: AC2239  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1526 <KUR>  
 A:Cross-references: GB:BA000019; PIDN:BA75165.1; PID:gl7132599; GSPDB:GN00179  
 A:Experimental source: strain PCC 7120  
 C:Genetics:  
 A:Gene: alr3466

Query Match 22.9%; Score 590; DB 2; Length 1526;  
 Best Local Similarity 34.1%; Pred. No. 2.3e-37;  
 Matches 132; Conservative 76; Mismatches 153; Indels 26; Gaps 8;  
 QY 103 CTSSMPGHAEEAVVSLNFSPDGAHLASGSDTTVRLMDLNTETPHFTCTGCHKOWILCVSWA 162  
 DB 1066 CLYTLOGHTSCVRSVVSFSPDGMALASGDDQIVRLWDISSGNCILYLTQGYTSWVRLVFS 1125  
 QY 163 PDGRKLASGCKAGSIIIDWPTGQGRPLSGHKKHINCLAWEPYHRDPECRKASAGD 222  
 DB 1126 PNGVTLANGSSDQIVRLMD-ISSKKCLYTLQGTTHWVNAF-----SPDGTLASGSD 1179  
 QY 223 GDCRIWDVKLQCLMNIAGHTNATVAVRWGAG-LIYTSKQRTVKWRAADGILCTFFS 281

Db 1180 QTVRLMDISSKCLYLQGHSTSWNVVFNPDGDTLASSGSDQTVRLWEINSKCLCTFQ 1239  
QY 282 CHAHWNINIALSTD-YVLTGTPHPVKDRSKSHLSLSTELQESALKRYQ-----A 331  
Db 1240 GHTSWNVSVFNPDMGLASG---SDKTVRLMDISSK---CLHTFOGHTWNWSVA 1291  
QY 332 VCPDEVESLVSCSDNTLYLRNNONKCVERTMTHQNVVNDVRYSPDVKLTASAPDKSV 391  
Db 1292 FNPDG-SMLASGSDQTVRLWEISSKCLHTFQGHSTSWNVSVFSPDCTMLASGSDQTV 1350  
QY 332 RLNRASDGOYMATFRGHVQAVYTVVANSADSLIVSGKSDTLKAVSWQTKLAQELPCHA 451  
Db 1351 RLMSISSGECILYTFILGHTNWNVGSVIFSPDGAILLASGSDQTVRLMSISSGKCLTILQHN 1410  
QY 452 DEFGVDWAPDGSRSVASGGDKVILKW 478  
Db 1411 NWGSI VSPDGTLLASGSDQTVRLW 1437  
RESULT 7  
Ti18521  
beta transducin-like protein - Podospora anserina  
C:Species: Podospora anserina  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: Ti18521  
R:Saupé, S.; Turcq, B.; Begueret, J.  
Gene 162, 135-139, 1995  
A:Title: A gene responsible for vegetative incompatibility in the fungus Podospora anserina  
A:Reference number: 218944; MUID:96009891; PMID:7557402  
A:Accession: Ti18521  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1356 <SAU>  
A:Cross-references: EMBL:L28125; NID:g607002; PID:g607003; PIDN:AAA65775.1  
C:Gene: het-e1  
A:Introns: 761/3  
Query Match 22.6%; Score 581; DB 2; Length 1356;  
Best Local Similarity 27.2%; Pred. No. 1e-36;  
Matches 138; Conservative 81; Mismatches 182; Indels 106; Gaps 7;  
QY 13 HTIOARLYVTGEAGPPIDLPAGITVQOLGLICNALLKNEEATPY---LFFVGEDEIKS 69  
Db 763 HTIRGLIALVRDG-----YFALSYRMIEKAPLQAYLSALVAPTDSMIKK 810  
QY 70 LEDTLDLASVDTEVIDIVYQAVFQVPTVTRCTSSMPGHAEAVVSLNFPDGAHLASG 129  
Db 811 IFKKEPFWTISTISVEAEW-----NACTQTLEGGHSSVLSVAFSADGQVAVS 859  
QY 130 SGDTTVRLMDLNTETPHFTCTGCHKQWLCVSWAPDGKRLASGCKAGSIIDPPTGQOKG 189  
Db 860 SDDTKIKINDTASGTCTQTLGGHGSVWSVAFSPDRERVASGSDDKTIKINDAASGTCT- 918  
QY 190 RPLSGHKHKLINGLAWEYPYHR-----DKINDAASGTCTQTLGGHGSVLSVAFSPD 212  
Db 919 QTLGGHGRVQSVAFSPDQGVASGSDDKTIKINDAASGTCTQTLGGHGSVLSVAFSPD 978  
QY 213 CRKLASASGDGDCRIWDVKGOCILMNIAGHTNATVAVRWGAG-LIYTSSKDRVTKWRA 271  
Db 979 GQVAVSSGDKTIKINDTASGTCTQTLGGHGSVWSVAFSPDQGVASGSDDKTIKINDT 1038  
QY 272 ADGTLCTFSGHAHWNINIALSTDYVLTGTPHPVKDRSKSHLSLSTELQESALKRYQA 331  
Db 1039 ASGTCTQTLGGHGWVQSVVFPD-----DPE 1062  
QY 332 VCPDEVESLVSCSDNTLYLRNNONKCVERTMTHQNVVNDVRYSPDVKLTASAPDKSV 391  
Db 1063 -----GQVAVSSDHTIKINDAVSGTCTQTLGGHGSVWSVAFSPDQGVASISDGTI 1117  
QY 332 RLNRASDGOYMATFRGHVQAVYTVVANSADSLIVSGKSDTLKAVSWQTKLAQELPCHA 451  
Db 1118 KINDAASGTCTQTLGGHGWVHSVAFSPDQGVASGSDDKTIKINDAASGTCTQTLGGH 1177

QY 452 DEFGVDWAPDGSRSVASGGDKVILKW 478  
Db 1178 GWQSVAFSPDQGVASGSDDKTIK 1204

## RESULT 8

AE1866  
WD-40 repeat protein [imported] - Nostoc sp. (strain PCC 7120)  
C:Species: Nostoc sp.  
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 30-Jun-2002  
C:Accession: AE1866  
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Takazawa, M.; Sugimoto, M.; Yamada, M.; Yasuda, M.; Tal  
DNA Res. 8, 205-213, 2001  
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacter  
A:Reference number: AB1807; MUID:21595285; PMID:11759840  
A:Accession: AE1866  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1708 <KUR>  
A:Cross-references: GB:BA000019; PIDN:BA72436.1; PID:g17129823; GSPDB:GN00179  
A:Experimental source: strain PCC 7120  
C:Gene: g10478  
A:Gene: all0478

Query Match 20.9%; Score 538.5; DB 2; Length 1708;  
Best Local Similarity 32.3%; Pred. No. 2.8e-33;  
Matches 147; Conservative 77; Mismatches 168; Indels 63; Gaps 15;

QY 42 GLICNALLKNEEATPYLFFVGEDEIKKSLIEDTLDLASVDTEVIDIVYQAVFQVPTV 101  
Db 1137 GSLINTLSKHTNVVNSVNF-----SPDALLIASASODKTVKLNK-----VG 1178  
QY 102 RCTSSMPGHAEAVVSLNFPDGAHLASGSDTTVRLWD-----LNTETPHFTCTGHRQWV 156  
Db 1179 QLVTTILQHGVDVNNASFSFGSLIASGSDKTVKLSREGKLLN-----TLSGHNDV 1232  
QY 157 LCYSWAPDQKRLASCKAGSIIDPPTGQOKGRPL---SGHKHKLINGLAWEYPYHRDPEC 213  
Db 1233 LGIATPDQGLTASGVADKKILWN-----RQKLLTKWQGHDDAILGVAN-----SPKG 1282  
QY 214 RKLASAGDGCRIWDVKGOCILMNIAGHTNATVAVRWGAG-LIYTSSKDRVTKWRA 272  
Db 1283 ETATFASPDQTIKLNW-RQGNLLKTLGHTAGVATVTFSPNGETIGSASIDATLKLW-SF 1340  
QY 273 DGLICTFSGHAHWNINIALSTDYVLTGTPHPVKDRSKSHLSLSTELQESALKRYQAV 332  
Db 1341 QGLLGLTGLGHSNWNVSFSFSD-----GRIFASGSDK---FVTLNRWDEVLLRNPGRD 1392  
QY 333 CPDEVESLVSCSDNTLYLRNNQ-----NKCVERMTGHQNVVNDVRYSPDVKLIAS 384  
Db 1393 GNDWVTISFSFGDGETLAAASRDOTVKILSRHGKLLNTFKGTGTGIVGAVNSPRQMIAS 1452  
QY 385 ASFDKSVRLWRASDQYMATFRGHVQAVYTVVANSADSLIVSGKSDTLKAVSWQTKLA 444  
Db 1453 ASKDOTVKLNW-QDGKTLHTLQGHODAVLAVASDSDQVITASAGDKIVKWS-QGQQLL 1510  
QY 445 QELPHGHADEVGVWAPDGSRSVASGGDKVILWA 479  
Db 1511 HTLQGHDAVNWVSFSDPGKLLIASVSDDTTVKLWS 1545

## RESULT 9

AE1810

WD-40 repeat protein [imported] - Nostoc sp. (strain PCC 7120)

C:Species: Nostoc sp.  
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 712  
C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 30-Jun-2002  
C:Accession: AE1810  
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.  
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.;



```
Db 1311 TGHERITSVKPF-----SPDGKILASASDKTIKFWNTD-GKPLKTLIAAHNQVNSINFS 1364
QY 253 G-AGLIYTSSKDRVKWRAADGILCTRFSGHAHWNNI-----ALSTDYVLRWG 301
Db 1365 SDSKTLVSAGADSTMVKW-IDGTLIKTISGEQIRDTVTFSPDNKVIASASSDKTVIR 1423
QY 302 PFHPVKDRSKSHLSL-----TELOESALKRYQ-----A 331
Db 1424 QLVYQSKQSNVNSVSPDGKTFASAGWGNITIMQRETLAHSLSTIQKNQNIITVS 1483
QY 332 VCPDEVESIVSCDDNTLYLRNNQNKQVERMTGHQNVNDVKYSPDVKLTIASASFDKSV 391
Db 1484 YSPDG-KTIATASADNIIKLWDSQTLIKTLGHKDRITLSFHPDQNTIASSADKTYI 1542
QY 392 RLNRASDQYMATFRGHVQAVYIVAWSADSLIVSGSKDSTLKV-----435
Db 1543 KIWRVNDGQLLTGTHGNDVTSVNFSPDGQFLASGSDTNVYKIMQTDGRLIKNITHGL 1602
QY 436 -----WSVOTKKLAQELPGHADEVFGVWDWAPDGSRVASGG 470
Db 1603 AIASVKFSPDSHTLASASDNTIKLWQVTDGKLNNLNGHIDGVTSLSFSPDGELASGS 1662
QY 471 KDVIKLW 478
Db 1663 ADNTIKLW 1670

RESULT 12
AG1889
WD-40 repeat protein [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp.
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C:Accession: AG1889
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriiguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AG1889
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-934 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA72622.1; PID:g17130010; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all0664

Query Match 19.3%; Score 497; DB 2; Length 934;
Best Local Similarity 29.7%; Pred. No. 2.1e-30;
Matches 126; Conservative 88; Mismatches 140; Indels 70; Gaps 15;

QY 106 SMPGHAEVYVSLNFSPDGHAHLSGSDTTVRLWDINTETPHFTCTGKHQWILCVSNAPDG 165
Db 445 TLEGHKDKVNSITFSPDGLIATVGTWMDNMTKLWLDGKELR-TFRGHQMIWSVSPDG 503
QY 166 KRLASGCKAGSIIWDPETGQGRPLSGHKHINCLAWEPYHRDPECKLASASGDGDC 225
Db 504 KOIATASGDRTVKLWSLD-GKEL-QTLRGHONGVSVTF-----SPDGKLIATASGDRTV 556
QY 256 RINDVKLGOCMLNIAHGTNAVTVRWGGAG-----255
Db 557 KLNSKR-GOELETLYGHDAVNSVAFSPDGTSTIATAGNDKTAIKMLNSPNSIIVRGHED 615
QY 256 -----LIYTSKDRTVKWRADGIL-CRTFSGHAWNNIALSTD--YVLR 299
Db 616 EYDVLFPNGKYIATASNDKTAKLWSIVGDKLDELTFNCHGRVKNLSFSPGKIAT 675
QY 300 TGFPHVPKDRSKSHLSLSTELQESALKRYQAV-----CPDEVESIVSCDDNTLYLRN 354
Db 676 TS-----WDKTAKLWNLID-GTLQKTLTGHKDTVWSVNFSPDG-OLTIATASEDKTVKLW-N 727
QY 355 NONKCVERTGHQNVNDVKYSPDVKLIASASFDKSVRLWRASDGOYMATFRGHVQAVYT 414
```

```
Db 728 ROGELLKTLPRQSSVYNSAVFSPDGKLIATAGWDKTVKIW-SIDGRLOKTLTGHYSINS 786
QY 415 VAWSADSLIVSGSKDSTLKVSVOTKKLAQELPGHADEVFGVWDWAPDGSRVASGGDKV 474
Db 787 VTFSPDGKLIASASDNTVAKIWLNGKEL-RTLGRHKNVYVHNTFSPDGKLIATASDNT 845
QY 475 IKLW 478
Db 846 VKLW 849

RESULT 13
AG12493
WD-repeat protein [imported] - Nostoc sp. (strain PCC 7120) plasmid pCC7120alpha
C:Species: Nostoc sp.
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C:Accession: AG12493
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriiguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Ta
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacter
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AG12493
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1189 <KUR>
A:Cross-references: GB:BA000020; PIDN:BA78213.1; PID:g17135667; GSPDB:GN00180
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr7129
A:Gene: plasmid

Query Match 19.3%; Score 496.5; DB 2; Length 1189;
Best Local Similarity 29.3%; Pred. No. 3.2e-30;
Matches 137; Conservative 78; Mismatches 170; Indels 83; Gaps 11;

QY 72 DTLDLASVDITENVIDI-----VYQPO-----AVFKVRPVT-----101
Db 573 DGCQLATSDTSGVINWNVNGKQLENCEHNSMIWDAFVSAPVPLASCGQDHTIKLWN 632
QY 102 ----RCTSSMPCHAEVYVSLNFSPDGHAHLSGSDTTVRLWDINTETPHFTCTGKHQWVL 157
Db 633 TTTGECFNLHGHTSIVTSVAFSPGKLIASSSYDHSYKVMWDLTGECIQFLGHDCVW 692
QY 158 CVSWAPDGRHLASGCKAGSIIWDPETGQGRPLSGHKHINCLAWEPYHRDPECKLA 217
Db 693 SVVFPVPGOILATAGENDTIKLWELQSGCL-KTLQGHQHWVTIAF-----NSGRLA 746
QY 218 SASGSDGCRINDVKLGOCMLNIAHGTNAVTVRWGGA-GLIYTSKDRTVKWRADGIL 276
Db 747 SGSDQNVKLVMDIHTGKCVMTLQGTGVVTSVAFNPXNDNLLSGSYDQSVKVMWRTGRC 806
QY 277 CRTFSGHAWNNIALSTDYVLRTPGFHPVKDRSKSHLSL-----TE 319
Db 807 LDTLKHTNRIMSA-----FHP-----QGHLYVSGGDDHAAKINBELGTGQCIK 850
QY 320 ELOESALKRYQAVCPDEVESIVSCDDNTLYLW-----RNNQKCVYER-MTGHQNVY 370
Db 851 TFOGHSNATYTTIAHWEHSLASGHEDQTIKLWDLNLHSPKSNVNTVFPFRLQGHNRV 910
QY 371 NDVKYSPDVKLIASASFDKSVRLWRASDGOYMATFRGHVQAVYTVAWSADSLIVSGSKD 430
Db 911 FSVVFSSTQGLIASSADRTIKLWSPHTGQCLTLHGHSWVWAIASFSLDDKLLASGSYD 970
QY 431 STLKVMVSVOTKKLAQELPGHADEVFGVWDWAPDGSRVASGGDKVIKLW 478
Db 971 HTVKIWNVSSGQCLOTLQGHGSLVAFSCDGKTLFSSGVEKLVQW 1018

RESULT 14
AB2410
```

WD-repeat protein [imported] - Nostoc sp. (strain PCC 7120)

C:Species: Nostoc sp.  
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 30-Jun-2002  
C:Accession: AB2410  
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimizu, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.  
DNA Res. 8, 205-213, 2001  
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena PCC 7120  
A:Reference number: AB1807; MUID:21595285; PMID:11759840  
A:Accession: AB2410  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1551 <KUR>  
A:Cross-references: GB:BA000019; PIDN:BA076533.1; PID:g17132214; GSPDB:GN00179  
A:Experimental source: strain PCC 7120  
C:Genetics:  
A:Gene: alr3119

Query Match 18.9%; Score 486; DB 2; Length 1551;  
Best Local Similarity 29.4%; Pred. No. 3e-29;  
Matches 121; Conservative 75; Mismatches 128; Indels 88; Gaps 16;  
QY 106 SMPGHAAYVSLNFPDGAHLASGSDTIVRLMDLN-TETPHFTCTGKQWVLCVSNAPD 164  
DB 989 SLOGHEITWNSAFDPGKIYATASDRTARLWNSGQOLAKF--QHGQYVRSVFSFPD 1046  
QY 165 GKRLASGCKAGSIIDWPTGQKGRPLSGHK-----KHINCLA-----WE 205  
DB 1047 GKHIATAGDDHARLWS-FSGQQLVQ-FPGHQCTWCISFSPDGKGIATATAADRTVRLWN 1104  
QY 206 -----PYHRD-----PECKLASAGDGDRIWDVKGLOCLMNIAGHTNAVTA 248  
DB 1105 LKGLLVRFPGHQDQWVDSFSDSYIATASDCTSLRNL-AGEQITFRGHGVWS 1163  
QY 249 VHWGAG-LIYTSSKDRVKNRAADGILCRPFESHAHWNNIALSTDYVLTGPFHVK 307  
DB 1164 VRFSPNQYIATSSDRTARVWN-LNQQLAQFSGQDYVRSVFSFPD-----GRI----- 1213  
QY 308 DRSKSHLSLSTELQESALKRYQAVCPDEVESLVSCDDNTLYLRNNQNKVCERMTGCHQ 367  
DB 1214 -----IATASSDRTVRLWHLNKOQ-FSAFQGHQ 1240  
QY 368 NYVNDVYSPDKLIASAFKSVRLWRASDGOYMATFRGHVQAVYTVWASDRLIVSG 427  
DB 1241 STVRSVDFSPDGQVVTAAADRTVRLWNLK-GEELIQFLGHRGKWSVFSFDPGKIYAT 1299  
QY 428 SKDSTLKVSVQTKLAQELFCHADEVFGVDWAPDGSRVASGGKDKVILWA 479  
DB 1300 SSDRTVRLWDI-TGQLLOQFPGHQGTWVSFSPDQGHIAATASDLTTLWS 1350

## RESULT 15

AH2195  
hypothetical protein alr3119 [imported] - Nostoc sp. (strain PCC 7120)  
C:Species: Nostoc sp.  
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 30-Jun-2002  
C:Accession: AH2195  
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimizu, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.  
DNA Res. 8, 205-213, 2001  
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena PCC 7120  
A:Reference number: AB1807; MUID:21595285; PMID:11759840  
A:Accession: AH2195  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-676 <KUR>  
A:Cross-references: GB:BA000019; PIDN:BA074818.1; PID:g17132214; GSPDB:GN00179  
A:Experimental source: strain PCC 7120  
C:Genetics:  
A:Gene: alr3119

Query Match 18.5%; Score 476.5; DB 2; Length 676;  
Best Local Similarity 27.6%; Pred. No. 5.2e-29;  
Matches 108; Conservative 72; Mismatches 119; Indels 92; Gaps 6;  
QY 91 POAVFKVRPVRCTSSMPGHAEAVVSLNFPDGAHLASGSDTIVRLMDLNTETPHFTCT 150  
DB 371 POLITRFEISQPYTLGASHASOVNSVAFSPNGEFLASGSDDKTIKWNLNKKQKIHTLP 430  
QY 151 GHKQWVLCVSNAPDGRKLASGCKRAGSIIDWPTGQKGRPLSGHKKHINCLAWPEYHRD 210  
DB 431 GHSQWVAIAFSPDKTILAS-----TG----- 452  
QY 211 PECKLASAGDGDRIWDVKGLOCLMNTAGHTNAVTAWRGGAG-LIYTSSKDRVKNW 269  
DB 453 -----ADRTIKLWNLATGREIRHLKHSQGVSAFSPDGKTLASGLUKTIKWL 502  
QY 270 RAADGILCRFESGHAHWNNIALSTDYVLTGPFHVKDRSKSHLSLSTELQESALKRY 329  
DB 503 NPATGKEIRTLQESHSQGVANVAFSPD----- 528  
QY 330 QAVCPDEVESLVSCDDNTLYLRNNQNKVCERMTGHNQVNDVYSPDKLIASAFDK 389  
DB 529 -----GKTLASGSDKTIKLNLTTSKVIHTLKGHSOLVMSVAFNSDSQTLASGSKD 581  
QY 390 SVRLWRASDGOYMATFRGHVQAVYTVW-SADSRLIVSGSKDSTLKVSVQTKLAQELP 448  
DB 582 TIKLWNLSTGKTIRTLRHSKDKVNSVAVYPRDSTVLASGSDNTIKLWNLTTGELIIRLTK 641  
QY 449 GHADVFVGDWAPDGSRVASGGK-DKVILWA 478  
DB 642 RDSGYIYVISPDRNLASGSAENIIKIW 672

Search completed: June 2, 2003, 07:02:33

Job time: 48 secs





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 2, 2003, 06:52:40 ; Search time 35 Seconds  
(without alignments)  
568.816 Million cell updates/sec

Title: US-09-830-980-1  
Perfect score: 2574  
Sequence: 1 MQETDTQEATPHITQARLV.....PDGSRVASGGKDKVILWAY 480

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	1092	42.4	515	1 YCW2_YEAST	P25382 saccharomyc
2	599	23.3	1258	1 YS00_ANASP	Q8Y3C2 anabaena sp
3	597	23.2	1693	1 YL24_ANASP	Q8Y377 anabaena sp
4	590	22.6	1526	1 YL46_ANASP	Q8Y311 anabaena sp
5	581	22.6	1356	1 HET1_PODAN	Q00808 podospira a
6	433	16.8	1693	1 Y163_SYNY3	Q55363 synecocyst
7	419	16.3	334	1 WDR5_HUMAN	Q9U9P9 homo sapien
8	415	16.1	361	1 WDS_DROME	Q9VJ38 drosophila
9	409.5	15.9	742	1 PKWA_THECU	P49695 thermomonos
10	372	14.5	376	1 YKY4_CAEEL	Q17963 caenorhabdi
11	371	14.4	577	1 TBL1_HUMAN	O60907 homo sapien
12	361.5	14.0	501	1 YH92_CAEEL	Q23256 caenorhabdi
13	359	13.9	614	1 TUL1_SCHPO	Q09715 schizosacch
14	354	13.8	409	1 LIS1_BOVIN	P43033 bos taurus
15	354	13.8	409	1 LIS1_HUMAN	P43034 homo sapien
16	354	13.8	409	1 LIS1_MOUSE	P43035 mus musculu
17	347.5	13.5	395	1 YZLL_CAEEL	Q93847 caenorhabdi
18	346	13.4	678	1 SCOB_EMENI	Q00659 emericella
19	320	12.4	423	1 WDRC_HUMAN	Q92417 homo sapien
20	319.5	12.4	542	1 FW1B_HUMAN	Q9UKB1 homo sapien
21	319	12.4	312	1 GLP1_LEIMA	Q25306 leishmania
22	318.5	12.4	586	1 TUL2_SCHPO	Q9U9G8 schizosacch
23	316	12.3	423	1 WDRC_MOUSE	Q9J1A4 mus musculu
24	315	12.2	312	1 GLP1_LEICH	Q27434 leishmania
25	314.5	12.2	518	1 TRCB_XENLA	Q91854 xenopus lae
26	313.5	12.2	605	1 FW1A_HUMAN	Q9Y397 homo sapien
27	312.5	12.1	605	1 POF1_SCHPO	Q87053 schizosacch
28	312.5	12.1	665	1 LIZ3_CAEEL	Q93990 caenorhabdi
29	309	12.0	640	1 WDR1_YEAST	P39014 saccharomyc
30	305	11.8	608	1 WDR1_DROME	Q9VU68 drosophila
31	303	11.8	514	1 TUP1_CANAL	P56093 candida alb
32	303	11.8	682	1 TUP1_KLULA	P56094 kluyveromyc
33	303	11.8	1248	1 APAF_HUMAN	O14727 homo sapien

09epv5 rattus norv  
p93107 chlamydomon  
p78706 neurospora  
088879 mus musculu  
Q93794 caenorhabdi  
P17343 caenorhabdi  
Q93277 gallus gall  
P16649 saccharomyc  
P16649 saccharomyc  
Q9NYS7 homo sapien  
P49846 drosophila  
Q08706 lymnaea sta  
Q01277 neurospora

## ALIGNMENTS

RESULT 1

ID	YCW2_YEAST	STANDARD;	PRT;	515 AA.
AC	P25382;			
DT	01-MAY-1992 (Rel. 22, Created)			
DT	01-FEB-1996 (Rel. 33, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Hypothetical 57.0 kDa Trp-Asp repeats containing protein in CPR4-SSK22			
DE	intergenic region.			
GN	YCR072C OR YCR72C.			
OS	Saccharomyces cerevisiae (Baker's yeast).			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
OC	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.			
OX	NCBI_TaxID=4932;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Ballesta J.P.G., Franco L., Hoenicka J., Jimenez A., Remacha M.,			
RA	Sanz E.;			
PL	Submitted (MAR-1992) to the EMBL/GenBank/DBJ databases.			
RN	[2]			
RP	REVISIONS.			
RA	Jimenez A.; (DEC-1992) to the EMBL/GenBank/DBJ databases.			
PL	Submitted (MAY-1992) to the EMBL/GenBank/DBJ databases.			
CC	-1- SIMILARITY: CONTAINS 8 WD REPEATS (TRP-ASP DOMAINS).			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	EMBL; X59720; CAA42270.1; -			
DR	PIR; S19487; S19487.			
DR	SGD; S0000668; YCR072C.			
DR	InterPro; IPR001680; WD40.			
DR	Pfam; PF00400; WD40; 8.			
DR	PRINTS; PR00320; GPROTEINBRPT.			
DR	ProDom; PD000018; WD40; 6.			
DR	SMART; SM00320; WD40; 8.			
DR	PROSITE; PS00678; WD.REPEATS.1; 5.			
DR	PROSITE; PS50082; WD.REPEATS.2; 7.			
DR	PROSITE; PS50294; WD.REPEATS.REGION; 1.			
KW	Hypothetical protein; Repeat; WD repeat.			
FT	REPEAT 141 172 WD 1.			
FT	REPEAT 184 214 WD 2.			
FT	REPEAT 227 264 WD 3.			
FT	REPEAT 276 305 WD 4.			
FT	REPEAT 318 387 WD 5.			
FT	REPEAT 400 430 WD 6.			
FT	REPEAT 442 472 WD 7.			
FT	REPEAT 484 514 WD 8.			
SQ	SEQUENCE 515 AA; 57040 MW; DFE603B79BF530A CRC64;			

Query Match 42.4%; Score 1092; DB 1; Length 515;

DR	EMBL: AP003590; BAB74499.1; -	DR	InterPro: IPR000767; Disease_resist.	DR	InterPro: IPR001680; WD40.	DR	Pham: PF00400; WD40; 14.	DR	PRINTS: PRO0364; DISEASERISIT.	DR	PRINTS: PRO0320; GPROTEINBRT.	DR	PRODOM: PD000018; WD40; 1.	DR	SMART: SM00320; WD40; 14.	DR	PROSITE: PS00678; WD_REPEATS_1; 9.	DR	PROSITE: PS0082; WD_REPEATS_2; 14.	DR	PROSITE: PS0294; WD_REPEATS_REGION; 1.	DR	Hypothetical protein; Repeat; WD repeat; Complete proteome.	FW	REPEAT 55 93	FT	REPEAT 640 679	FT	REPEAT 682 721	FT	REPEAT 724 763	FT	REPEAT 766 807	FT	REPEAT 809 849	FT	REPEAT 850 889	FT	REPEAT 892 931	FT	REPEAT 976 1017	FT	REPEAT 1019 1059	FT	REPEAT 1060 1101	FT	REPEAT 1103 1143	FT	REPEAT 1144 1183	FT	REPEAT 1186 1227	SQ	SEQUENCE 1258 AA; 139513 MW; 45DF03B91170C451 CRC64;	Query Match 23.3%; Score 599; DB 1; Length 1258;	Best Local Similarity 32.7%; Pred. No. 7.5e-39;	Matches 131; Conservative 66; Mismatches 150; Indels 54; Gaps 7;	QY	103	CTSSMPGHAENVSLNPSDPGAHLASSGDTTVRLWDLNTEPHCTGCHKQWLVCSHA	162	718	CKTKTTHGEHEVFSVAFHPDGETLASAGDKTIKLDIQDGCLOQTTHGHDWRCVAFS	777	163	PDGRLASGGKAGSLIWDPTGQKGRPLSGHKKHINCLAWPEYHDPCKRLASASGD	222	778	PDGNTLASSAADHTIKLDVSGQKCL-RTLKSHTGWRSVAF-----SADQTLASSGD	831	223	GDRCITWDVKLQCLMNIAGHTNAVTAVRWG-GAGLIYTSKSDRTVKMWRADGILCRWFS	281	832	RTIKWNTHTGECLTYIIGHTNSVTSIAYSPDSKILVSGSDRTIKLWDCQTHICIKTLH	891	282	GHAHWNIALSTD-----YVLRTGPFHPVKDRSKLSLSTEELOESAL	326	892	GHTNEVCSVAFSPDGTCLACVSLDQSVRLNCRGT-----QCL	929	327	KRYQ-----AVCPDEVESLVSCDDNTLYLRNNONKCVEMTGHONVNDVKYSP	377	930	KAWYNTDWPALVAFSPDR-QILASGSDTKVLWDQWTKYISLEGHDTFYTGIAFSP	988	378	DYKLTLASAFKSVRLWRASDGOYMAFTRGHVQAVTVTAWSADSLIVSGSKDSTLKVWS	437	989	DSQTLASATSSVRLWNISTGQCQLLEHTDWYAVVFHPQGIATGSDCTVKLWN	1048	438	VOTKKLAQELPGHADEVFGVDWADPGSEVASGDKVIKLM	478	1049	ISTGQCLTKLSEHSKILGNWSPDQLLASADQSVRLM	1089	RESULT 3	YL24_ANASP	STANDARD;	PRT; 1683 AA.	AC	QBY57;	DT	15-JUN-2002 (Rel. 41, Created)	DT	15-JUN-2002 (Rel. 41, Last sequence update)	DT	15-JUN-2002 (Rel. 41, Last annotation update)	DE	Hypothetical WD-repeat protein all2124.	GN	ALL2124.	OS	Anabaena sp. (strain PCC 7120).	OC	Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
----	-------------------------------	----	--------------------------------------	----	----------------------------	----	--------------------------	----	--------------------------------	----	-------------------------------	----	----------------------------	----	---------------------------	----	------------------------------------	----	------------------------------------	----	--	----	---	----	--------------	----	----------------	----	----------------	----	----------------	----	----------------	----	----------------	----	----------------	----	----------------	----	-----------------	----	------------------	----	------------------	----	------------------	----	------------------	----	------------------	----	--	--	---	--	----	-----	---	-----	-----	--	-----	-----	--	-----	-----	--	-----	-----	---	-----	-----	--	-----	-----	--	-----	-----	--	-----	-----	--	-----	-----	---	-----	-----	--	-----	-----	--	------	-----	---	-----	------	--------------------------------------	------	----------	------------	-----------	---------------	----	--------	----	--------------------------------	----	---	----	---	----	---	----	----------	----	---------------------------------	----	---

```

OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriiguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.,
RA *Complete genomic sequence of the filamentous nitrogen-fixing
RA Cyanobacterium Anabaena sp. strain PCC 7120.*;
RA DNA Res. 8:205-213(2001).
RL -!- SIMILARITY: CONTAINS 13 WD REPEATS (TRP-ASP DOMAINS).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AP003588; BAB73823.1;
CC InterPro; IPR001680; WD40.
CC Pfam; PF00400; WD40; 14.
CC PRINTS; PR00320; GPROTEINRPT.
CC SMART; SM00320; WD40; 14.
CC PROSITE; PS00678; WD_REPEATS.1; 3.
CC PROSITE; PS50082; WD_REPEATS.2; 14.
CC PROSITE; PS50294; WD_REPEATS_REGION; 1.
CC KW Hypothetical protein; Repeat; WD repeat; Complete proteome.
FT REPEAT 1070 1109
FT REPEAT 1111 1150
FT REPEAT 1153 1192
FT REPEAT 1195 1236
FT REPEAT 1280 1318
FT REPEAT 1320 1359
FT REPEAT 1361 1400
FT REPEAT 1402 1441
FT REPEAT 1443 1483
FT REPEAT 1486 1525
FT REPEAT 1528 1567
FT REPEAT 1570 1609
FT REPEAT 1611 1650
SQ SEQUENCE 1663 AA; 185459 MW; F7CB361FF54F7137 CRC64;

Query Match 23.2%; Score 597; DB 1; Length 1693;
Best Local Similarity 26.2%; Pred. No. 1.6e-38;
Matches 149; Conservative 91; Mismatches 174; Indels 154; Gaps 9;

QY 36 ITTQQLGLCNALLKNEATPYLFFVGDEDEIKKSLDTDLASVDTENVIDIVYQPAVF 95
DB 1019 LSHQQLAALIALSKAAQOVNHI-----AVPNHLKLATVTT-----LQQALF 1060

QY 96 KVRPVTRCTSSMPGHAEAVVSLNFPDGAHLASGSDPTVRLMDLNTPEPHTCTGKHQW 155
DB 1061 EMQERNR----LEGHKDGVISISISIRSDGQTIIASGSLDTIKLSRDGRL--FTLNGHEDA 1115

QY 156 VILCVSWAPDCKRLASGCKAGSIIIDWPETGQOKGRPLSGHKHINCLAWEPYHRDPECK 215
DB 1116 VYSVSFSDPQQTIIASGSDKTIKLWQTSDDL--KTITGHEQIVNVV-----YFSPDGN 1169

QY 216 LASASGDGCRIVDWKVLGOCMLNAGHTNAYTVAPWGAG--LYITSKDRYVKKWRAADG 274
DB 1170 LASASSDHSIKLWDITSQLLMLTGHSAQVITVRFSPDQQTAAAGSDKIVKLMHRQDG 1229

QY 275 ILCTFSGHAWVNNIALSTD----- 295
DB 1230 KLLATLNGHODVWVNSLSPDQGTILASASADKTIKLMRIAGKLVKTLKGHNDSVWDVNF 1289

QY 296 -----YVLRTGPFPHVKDRSKSHLSLSTEEQLQESAK 327
DB 1290 SSDGKATASARDNTIKLWNRHGHTELETFTGHSGGVAVNFPDPSNIIASASLQNTIRLW 1349

```

## RESULT 4

```

YY46_ANASP
ID YY46_ANASP STANDARD; PRT; 1526 AA.
AC Q8YR11;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Hypothetical WD-repeat protein alr3466.
GN ALR3466.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriiguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.,
RA *Complete genomic sequence of the filamentous nitrogen-fixing
RA Cyanobacterium Anabaena sp. strain PCC 7120.*;
RA DNA Res. 8:205-213(2001).
RL -!- SIMILARITY: CONTAINS 16 WD REPEATS (TRP-ASP DOMAINS).
CC -!- SIMILARITY: CONTAINS 1 PENTAPEPTIDE REPEAT DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AP003593; BAB75165.1;
CC InterPro; IPR001646; 5peptide_repeat.
CC InterPro; IPR001680; WD40.
CC Pfam; PF00805; Pentapeptide; 1.
CC Pfam; PF00400; WD40; 15.
CC PRINTS; PR00320; GPROTEINRPT.
CC PRODOM; PD000018; WD40; 8.
CC SMART; SM00320; WD40; 15.
CC PROSITE; PS00678; WD_REPEATS.1; 11.
CC PROSITE; PS50082; WD_REPEATS.2; 15.
CC PROSITE; PS50294; WD_REPEATS_REGION; 1.
CC KW Hypothetical protein; Repeat; WD repeat; Complete proteome.
FT REPEAT 334 376
FT DOMAIN 823 862
FT REPEAT 862 901
FT REPEAT 904 945
FT REPEAT 946 985
FT REPEAT 988 1027
FT REPEAT 1030 1069
FT REPEAT 1072 1111
FT REPEAT 1114 1153

```

```
FT REPEAT 1156 1195 WD 9.
FT REPEAT 1198 1237 WD 10.
FT REPEAT 1240 1279 WD 11.
FT REPEAT 1282 1321 WD 12.
FT REPEAT 1324 1363 WD 13.
FT REPEAT 1366 1405 WD 14.
FT REPEAT 1408 1447 WD 15.
FT REPEAT 1450 1491 WD 16.
SQ SEQUENCE 1526 AA; 170376 MW; D5DF272509B4A738 CRC64;

Query Match 22.98; Score 590; DB 1; Length 1526;
Best Local Similarity 34.18; Pred. No. 4.9e-38;
Matches 132; Conservative 76; Mismatches 153; Indels 26; Gaps 8;

QY 103 CTSSMPGHAEEAVSLNFPDGAHLASGSDTTVRLMDLNTETPHFTCTGKHQVLCVSA 162
DB 1066 CLYTLOGHTSCVRSVSPDGAHLASGDDQIVRLNDISSGNCLYTLQGYTSWVRVLS 1125
QY 163 PDGKRLASGCKAGSIITWDPEGQOKGRPLSGHKHKLAWEPYHRDPCKRLASGSD 222
DB 1126 PNGVTILANGSSDQIVRLMD-ISKKCLYTLQHTNWNVAFAF-----SPDGAHLASGSD 1179
QY 223 GDCRIWDKLGOCCLMNTAGHTNAVAVRWGAG-LIYTSKORTYKMRADGILCTFS 281
DB 1180 QTVRLMDISSKCLYLIQHTSWNSVFNPDGSTLASGSDQTVRLWEINSSKCLCTFQ 1239
QY 282 GHAWVNNIALSTD-VYLRTGPFPHVKDRSKSHLSLSTELAESALKRYQ-----A 331
DB 1240 GHTSWNSVFNPDGSMLASGS-----SDKTVRLMDISSSK-----CLHTFQHTNWNVSA 1291
QY 332 VCPDEVESLVSCDDNTLYLRNQNCKVERMTGHQNVNDVYKSPDVKLIASAFDQSV 391
DB 1292 FNDPG-SMLASGSDQTVRLWEISSKCLHTFQHTSWNSVTFSPDGTMLASGSDQTV 1350
QY 392 RLWRASDGOYMATFRGHVQAVYVWASDLSRLIVSGSKSTLKVNSVOTKKLAQELPGA 451
DB 1351 RLWISSGCECLYTLFHTNWNVSVIFPDGAILASGSDQTVRLWSISSGKCLVTLQGHN 1410
QY 452 DEYFGVDWAPDGSRSVASKGDKVILW 478
DB 1411 NWQSVIFSPDGTLLASGSDQTVRLW 1437

RESULT 5
HETL_PODAN STANDARD; PRT; 1356 AA.
AC Q00808;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Vegetative incompatibility protein HET-E-1.
GN HET-E1.
OS Podospora anserina.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Lasiosphaeriaceae; Podospora.
OX NCBI_TaxID=5145;
RN [1]
RP MEDLINE=96009891; PubMed=7557402;
RA Saue S., Turcq B., Bequeret J.;
RT "A gene responsible for vegetative incompatibility in the fungus
RT Podospora anserina encodes a protein with a GTP-binding motif and G
RT beta homologous domain.";
RL Gene 162:135-139(1995).
CC -!- FUNCTION: RESPONSIBLE FOR VEGETATIVE INCOMPATIBILITY THROUGH
CC SPECIFIC INTERACTIONS WITH DIFFERENT ALLELES OF THE UNLINKED GENE,
CC HET-C.
CC -!- SIMILARITY: CONTAINS 10 WD REPEATS (TRP-ASP DOMAINS).
CC -!- SIMILARITY: CONTAINS 1 NACHT DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
```

```
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L28125; AAA85775.1; -.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 10.
DR PRINTS; PR00320; GPROTEINRPT.
DR PRODOM; PD000018; WD40; 10.
DR SMART; SM00320; WD40; 10.
DR PROSITE; PS00837; NACHT; 1.
DR PROSITE; PS00678; WD_REPEATS_1; 10.
DR PROSITE; PS00082; WD_REPEATS_2; 10.
DR PROSITE; PS0294; WD_REPEATS_REGION; 1.
KW GTP-binding; Repeat; WD repeat.
FT DOMAIN 294 629
FT NP_BIND 300 307
FT REPEAT 839 869
FT REPEAT 881 911
FT REPEAT 923 953
FT REPEAT 965 995
FT REPEAT 1007 1037
FT REPEAT 1049 1079
FT REPEAT 1091 1121
FT REPEAT 1133 1163
FT REPEAT 1175 1205
FT REPEAT 1217 1247
SQ SEQUENCE 1356 AA; 149765 MW; 965FB319844E0651 CRC64;

Query Match 22.64; Score 581; DB 1; Length 1356;
Best Local Similarity 27.24; Pred. No. 2.1e-37;
Matches 138; Conservative 81; Mismatches 182; Indels 106; Gaps 7;

QY 13 HTIQARLVYVTEEGAGPPIDLPAGITTOQLGLICNALIKKEATPY---LFFVGEDEIKKS 69
DB 763 HTIRGLIAIVRDG-----YRFALSYRIITEKAPLQAVTSALVPAPTDMSIKK 810
QY 70 LEDTDLASVDTEENVIVYQPAVEKVRPVTRCTTSMPGHAEEAVSLNFPDGAHLASG 129
DB 811 IFKKEPGWISTISVVEAEW-----NACTQTELEGHSSVLSVAFSADGQRVASG 859
QY 130 SGGTVRLMDLNTETPHFTCTGKHQVLCVSWAPDGRKRLASGCKAGSIITWDPTGQKG 189
DB 860 SDDKTIKWDTASGCTGTQTELEGHGSVMSVAFSPDRERVASGDDKTIKIWDASGTC- 918
QY 190 RPLSGHKHKLAWEPYHR-----DPE 212
DB 919 QTELEGHGSVMSVAFSPDQGVASGSDHTTIKIWDASGCTGTQTELEGHSSVLSVAFSPD 978
QY 213 CRKLASGSDGDCRIMDWKLGOCCLMNTAGHTNAVAVRWGAG-LIYTSKDRIVKRWRA 271
DB 979 GQVAVSGSGDKTIKIWDASGCTGTQTELEGHGSVMSVAFSPDQGVASGDDKTIKIWD 1038
QY 272 ADGILCTFSGHAHWNNIALSTDYVLRTPFHPVKDRSKSHLSLSTELAESALKRYQA 331
DB 1039 ASGCTGTQTELEGHGSVMSVAFSPD-----1062
QY 332 VCPDEVESLVSCDDNTLYLRNQNCKVERMTGHQNVNDVYKSPDVKLIASAFDQSV 391
DB 1063 -----GQVAVSGSDHTTIKIWDASGCTGTQTELEGHGSVMSVAFSPDQGVASGSDGTI 1117
QY 392 RLWRASDGOYMATFRGHVQAVYVWASDLSRLIVSGSKSTLKVNSVOTKKLAQELPGA 451
DB 1118 KIWDASGCTGTQTELEGHGSVMSVAFSPDQGVASGSDGTI KIWDASGCTGTQTELEGH 1177
QY 452 DEYFGVDWAPDGSRSVASKGDKVILW 478
DB 1178 GWQSVAFSPDQGVASGSDKTIKI 1204

RESULT 6
Y163_SYNY3
```

ID Y163\_SUNY3 STANDARD; PRT: 1693 AA.  
AC Q55563;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 16-JUL-1998 (Rel. 36, Last sequence update)  
DE 16-OCT-2003 (Rel. 40, Last annotation update)  
DE Hypothetical WD-repeat protein slf0163.  
GN SLF0163.  
OS Synechocystis sp. (strain PCC 6803).  
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.  
OA NCBI\_TaxID=1148;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96127529; PubMed=8590279;  
RA Kaneko T., Tanaka A., Sato S., Kotani H., Suzuki T., Miyajima N.,  
RA Sugita M., Tabata S.;  
RT "Sequence analysis of the genome of the unicellular cyanobacterium  
RT Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb  
RT region from map positions 64% to 92% of the genome.";  
RL DNA Res. 2:153-166(1995).  
CC -1- SIMILARITY: CONTAINS 16 WD REPEATS (TRP-ASP DOMAINS).  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: D63999; BAA10064.1; -  
DR InterPro: IPR001680; WD40.  
DR Pfam: PF00400; WD40. 16.  
DR PRINTS: PR00320; GPROTEINREP.  
DR PRODOM: PD000018; WD40. 14.  
DR SMART: SM00320; WD40. 16.  
DR PROSITE: PS00678; WD\_REPEATS\_1; 8.  
DR PROSITE: PS50082; WD\_REPEATS\_2; 35.  
DR PROSITE: PS50294; WD\_REPEATS\_REGION; 1.  
KW Hypothetical protein, WD\_REPEATS\_REGION: 1.  
FT REPEAT 108 102 WD 1.  
FT REPEAT 109 103 WD 2.  
FT REPEAT 109 103 WD 3.  
FT REPEAT 109 103 WD 4.  
FT REPEAT 1135 1165 WD 5.  
FT REPEAT 1176 1206 WD 6.  
FT REPEAT 1217 1247 WD 7.  
FT REPEAT 1258 1288 WD 8.  
FT REPEAT 1299 1329 WD 9.  
FT REPEAT 1340 1370 WD 10.  
FT REPEAT 1381 1411 WD 11.  
FT REPEAT 1422 1452 WD 12.  
FT REPEAT 1463 1493 WD 13.  
FT REPEAT 1504 1534 WD 14.  
FT REPEAT 1545 1575 WD 15.  
FT REPEAT 1586 1616 WD 16.  
FT REPEAT 1627 1657 WD 16.  
SQ SEQUENCE 1693 AA; 189935 MW; 0977A827A0251CFF CRC64;  
Query Match 16.8%; Score 433; DB 1; Length 1693;  
Best Local Similarity 25.0%; Pred. No. 1.le-25;  
Matches 152; Conservative 82; Mismatches 203; Indels 172; Gaps 22;  
QY 13 HTIQARL-----VYTGEEAGPDPIDLPAGITTOOL-----GLICNALL 49  
DB 995 HSILARINECRCHHEGPPVTLRISPSMWTPLVLTA--TYNGIALWSFHEGLINLVR 1052  
QY 50 KNEATPYLFVGEDE--IKKSLEOTLDLASVDTEVT-----DIVQPOAVFKV 97  
DB 1053 GQCEATLWDSADGQYFATASADHTVLMQWREGVEATLGHEDWVRSVHFSPIHQPLV 1112  
QY 98 -----RPVTRCTSSMPGAEAVVSLNFSPDGAHLASGSGTIVTLWLDNT 142  
DB 1113 TSGQDNTARIWNFAGEQLTLC-----OGHADWVNAEFCNCHQIILLTASRDGTARLWDLG 1168

QY 143 ETPHPTCTGHKQWVLCVSWADPGKRRLKASCKAGSIIIWDPETGQOKRPLSGHKHINCL 202  
DB 1169 REIGL-COGHTSWVRNAQFSPDGQWIVTCSADGTARLWD--LSSQCFVAVLKQHWYRNA 1225  
QY 203 AWEPYHRDPECKLASASGDGCRIMDVKLGOCLMNIAGHTNVAIVRGGAG-LIYTS 261  
DB 1226 LW-----SPDGHIIITSSSDGTARVWS-RHGKCLGTLRGHHDHNGARFSLGQKIVTYS 1279  
QY 262 KDRVTKWNR-----AADGILCRFTS----- 281  
DB 1280 TDTARLWTKEGTLTLIRGHQKEVDYDAFDSADGFEVTVTSADQTAQWDISQDKTIILT 1339  
QY 282 GHAWVNNIALSTDYVLRTPFPFVKD-----RSKSLSLSTEEQLSALKRYQA----- 331  
DB 1340 GSHHWVRN-----AHFNPKGDRLLTVSRDKTARLWTTGECVAVLADHGWYRE 1388  
QY 332 --VCPDEVESLASCSDNTLYLRNNONKCVERTGCHQVYNDVYKSPDVKLIASAFDK 389  
DB 1389 GQFSPDG-QWIVTGSADKTAQLW-NVLGKRLTVLRGHQDAVLNVRFSPDSQIVTASDKG 1446  
QY 390 SVRLWR-----ASDQYMAT-----FRGHV 409  
DB 1447 TARVNNNTGRELAVLRHYEKNIFAEEFADGQFIVTASDNTAGIWEIVGREVGICRGHE 1506  
QY 410 QAVTVVWASDRLVSGSKDSTLKWVSVQTKLAQELPGHAEVFCVDWAPDGRVRSAG 469  
DB 1507 GPVYFAQFSADSRYLITLTSVDNTARIWDFLGRPLL-TLAGHQSVIVQARESPGGLIATV 1565  
QY 470 GKDKVVKLW 478  
DB 1566 SADHTARLW 1574  
RESULT 7  
WDNS\_HUMAN STANDARD; PRT: 334 AA.  
AC Q5UGP9; Q5NW7; Q5IVAS;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE WD-repeat protein 5 (WD repeat protein BIG-3).  
GN WDNS OR BIG.  
OS Homo sapiens (Human), and  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606, 10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC SPECIES-Human; TISSUE-Uterus;  
RA Young J.M., Woodward K.J., Aziz S., Burley M., Kwiatkowski D.J.,  
RA Povey S.;  
RT "Cloning of a sugar transporter gene, a G-beta subunit like gene and  
RT three novel genes in human chromosome 9q34.";  
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC SPECIES-Human;  
RA Watanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,  
RA Suzuki Y., Oobayashi M., Nishi T., Shibahara T., Tanaka T.,  
RA Nakamura Y., Isoigai T., Sugano S.;  
RT "NEO human cDNA sequencing project."  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC SPECIES-Human; TISSUE-Lung;  
RA Strausberg R.;  
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RC SPECIES-Mouse; STRAIN-C57BL/6J;  
RX MEDLINE=21588136; PubMed=11551928;  
RA Gori F., Divieti P., Demay M.B.;



```
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spadling A.C., Stapleton M., Strong R., Sun E.,
RA Svrlkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-P., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.:
RA "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000)
CC -1- SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AF233288; AAF43418.1; -
DR EMBL; ALI38972; CAB7292.1; -
DR EMBL; AE003424; AAF45791.1; -
DR FlyBase; FBgn0040066; wds.
DR InterPro; IPR001632; Gprotein_B.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 7.
DR PRINTS; PR00319; GPROTEINB.
DR PRINTS; PR00320; GPROTEINBPT.
DR PRODOM; PD000018; WD40; 6.
DR SMART; SM00320; WD40; 7.
DR PROSITE; PS00678; WD REPEATS 1; 4.
DR PROSITE; PS0082; WD REPEATS 2; 6.
DR PROSITE; PS0294; WD REPEATS_REGION; 1.
KW Repeat; WD repeat.
FT REPEAT 62 100 WD 1.
FT REPEAT 104 142 WD 2.
FT REPEAT 146 184 WD 3.
FT REPEAT 188 226 WD 4.
FT REPEAT 230 269 WD 5.
FT REPEAT 273 314 WD 6.
FT REPEAT 318 358 WD 7.
SQ SEQUENCE 361 AA; 39041 MW; B2A64127F1E0292D CRC64;

Query Match
Best Local Similarity 16.1%; Score 415; DB 1; Length 361;
Matches 103; Conservative 64; Mismatches 127; Indels 94; Gaps 7;

OY 97 VRPVTRCTSSMPGHAENVSLNFPDGAHLASGSGDTTVRLMDLNTETPHCTGKHQWV 156
DB 58 VRENTTLKFTLAGTKAVSAVKEPSPNGEGLASSADKLKIKWGYDGGKFKETISGHKLG 117
OY 157 LCVSNAADPKRLASGCKAGSIILWDPETGQGRPLSGHKKHINCILAWEPYHRDPCKRL 216
DB 118 SDVANSDDSLVSGSDDKTLKVELSTGKSL-KTLKGHSNYVFCNF-----NPQSNLI 171
OY 217 ASASGDGCRIDWVKLGQCLMNIAGHTNAVTVRWGAG-LIYTSKDRVTKMRAADGI 275
DB 172 VSGSFDESVRIMDVRTGCKLTPAHSDPVSAVFNKRDGSLIVSSYDGLCRIDWTFASQ 231
OY 276 LCRTFSGAHWNNIALSTDYVLRTPGPHPKDRSKSHLSLSTELQESALKRYQAVCPD 335
DB 232 CLKTL-----
OY 336 EVESILVSCDDNTLYLRNNQKNCVMTGHQNVNDVKYSPDKLIASAPKSVLRWR 395
DB 237 -----IDNDNP-----PVSFVAFSPNGKYLALILDTILKLD 269
OY 396 ASDQGYMATFGHVAQVYIVAMS---ADSLIVSGSKDSTLKWVSTQKKLAQELPGHD 452
DB 270 YSKGCKLTYTGHNEKCYIFANFSVTGCKWIVSGSDNMTYINWLSQKSVQKQGHRTD 329
```

```
OY 453 EYFGVDWAPDGSRVASGG--KDKVIKILW 478
DB 330 TVLTACHPTENIIASAALENDKTIKILW 357

RESULT 9
PKWA_THRCU
TD PKWA_THRCU STANDARD; PRT; 742 AA.
AC P49695;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Putative serine/threonine-protein kinase pkwa (EC 2.7.1.-).
GN PKWA OR PKWI.
OS Thermomopora curvata.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetaceae;
OC Actinomycetales; Streptosporangineae; Thermomoporaaceae;
OC Thermomopora.
OX NCBI_TaxID=2020;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CCM 3352;
RX MEDLINE=96200125; PubMed=8631732;
RA Janda L., Tichy P., Spizek J., Petricek M.;
RT "A deduced Thermomopora curvata protein containing
RT serine/threonine protein kinase and WD-repeat domains.";
RL J. Bacteriol. 178:1487-1489(1996).
CC -1- FUNCTION: MAY PLAY A REGULATORY ROLE DURING THE COMPLEX GROWTH
CC CYCLE AND IN SECONDARY METABOLITE PRODUCTION.
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC -1- SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AF115313; AAB05822.1; -
DR InterPro; IPR000719; Ser_pkinase.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00400; WD40; 7.
DR PRINTS; PR00320; GPROTEINBPT.
DR PRODOM; PD000001; Euk_pkinase; 1.
DR SMART; SM00320; WD40; 7.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE; PS00678; WD_REPEATS_1; 5.
DR PROSITE; PS0082; WD_REPEATS_2; 7.
DR PROSITE; PS0294; WD_REPEATS_REGION; 1.
KW Transferase; Serine/threonine-protein kinase; ATP-binding; Repeat;
KW WD repeat.
DB DOMAIN 16 266 PROTEIN KINASE.
FT NP_BIND 22 30 ATP (BY SIMILARITY).
FT BINDING 44 44 ATP (BY SIMILARITY).
FT ACT_SITE 138 138 BY SIMILARITY.
FT DOMAIN 301 394 PRO/GLU/ASP-RICH (SPACER).
FT REPEAT 455 496 WD 1.
FT REPEAT 497 538 WD 2.
FT REPEAT 539 580 WD 3.
FT REPEAT 581 621 WD 4.
FT REPEAT 622 663 WD 5.
FT REPEAT 664 705 WD 6.
FT REPEAT 706 742 WD 7.
SQ SEQUENCE 742 AA; 78950 MW; AC1734640DB4383D CRC64;

Query Match
15.9%; Score 409.5; DB 1; Length 742;
```





```

CC EMBL; Y12781; CAA73319.1; -.
DR Genew; HGNC:11585; TBLX.
DR MIM; 300196; -.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 8.
DR PRINTS; PR00320; GPROTEINBRPT.
DR ProDom; PD000018; WD40; 5.
DR SMART; SM00320; WD40; 8.
DR PROSITE; PS00678; WD_REPEATS_1; 4.
DR PROSITE; PS50082; WD_REPEATS_2; 6.
DR PROSITE; PS50294; WD_REPEATS_REGION; 1.
KW Repeat; WD repeat.
FT REPEAT 230..269 WD 1.
FT REPEAT 286..325 WD 2.
FT REPEAT 327..366 WD 3.
FT REPEAT 360..409 WD 4.
FT REPEAT 400..449 WD 5.
FT REPEAT 452..500 WD 6.
FT REPEAT 503..542 WD 7.
FT REPEAT 544..576 WD 8.
FT DOMAIN 159..175 POLY-ALA.
SQ SEQUENCE 577 AA; 62495 MW; D830A37781E2A15C CRC64;

Query Match 14.4%; Score 371; DB 1; Length 577;
Best Local Similarity 25.8%; Pred. No. 1.8e-21;
Matches 128; Conservative 52; Mismatches 213; Indels 104; Gaps 14;

QY 6 TQEQATPHITQARLYVTGEEAGPPIDLPAGITTOQLGLICNALLKNEEATPYLFFVGEDE 65
DB 145 TQEQAFREKLAQQOASAAAAATAATAATTAGVSQHPNSKNREAT-----VNGEEN 200
QY 66 IKKSLEDTLDLASVDENVIDIVYQPAQVFKVRPTVTRCTSSMPGHAENVSLNFSPDGAH 125
DB 201 RAHSVNNHAKPWEIDGE--VEI-----PSSKAT-VLRGHESEVFCANPVSOL 246
QY 126 LASGSGDVTVRDLNLTETPHCTCTGKQWVL--CV-----SWAPDGR 167
DB 247 LASGSGDSTARINLWNSN-----GGSTQLVLRHCIREGGHDVPSNKNVSLDWTNGTL 302
QY 168 LASGKAGSIIIDWPEGQKGRPLSGHKKHINCLAWEPYHRDPCBK-----LASASGDD 224
DB 303 LATGSYDGFARIW--TEDG-NLASTLGHQKGFALAWN-----RAGNYLSAGVKT 352
QY 225 CRTDVKLGCLMNIAGHTNATVNRGAGGLIYTSKDRTPKVMRAADGILCRTFSGHA 284
DB 353 TIIDWAGTGAQKQOFFHSAPALDWNQNTTFASCTDMCIHVCRCORPVKTEQGH 412
QY 285 HWNNIALSTDYVLRGPHFPVDRSKSHLSLSTEEQLQESALKRYQAVCPDEVESLVSCS 344
DB 413 NEYNALKWDPSGML-----LASCS 431
QY 345 DNTLYLRNNQNKVERMTGHONVNDVKYSPDVK-----LIASAFDKSYRLWR 395
DB 432 DMTLKIKMSQEQVCHDLOAHNKEIYTIKWSPTGATSNPNINILASAFDSTVRLWD 491
QY 396 ASDGYMATFRGHVQAVYTVAMSDRLISVSGSKDLKWSVQTKLAQELPGHAEVDF 455
DB 492 IERGCTHTLTKHOEPYISVAFSPDGKYLASGDFRCVHIWNTOSGNLVHSYRG-TGGIF 550
QY 456 GYDWAPEGDSRVASGGKD 472
DB 551 EVCWNARGDKVGASAD 567

RESULT 12
YH92.CABEL
ID YH92.CABEL STANDARD; PRT; 501 AA.
AC Q23256;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Hypothetical 34.5 kDa Trp-Asp repeats containing protein ZC302.2 in

```

```

DE chromosome V.
OS ZC302.2.
GN Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBL_TaxID=6239;
RN 111
RP SEQUENCE FROM N.A.
RC SPRAIN-Bristol NZ;
RA Kelley E.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).
CC -!- SIMILARITY: STRONG, TO K04G11.4 AND C14B1.4.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; Z73978; CAA98293.1; -.
DR WormPep; ZC302.2; CE06574.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 7.
DR PRINTS; PR00320; GPROTEINBRPT.
DR ProDom; PD000018; WD40; 4.
DR SMART; SM00320; WD40; 7.
DR PROSITE; PS00678; WD_REPEATS_1; 3.
DR PROSITE; PS50082; WD_REPEATS_2; 6.
DR PROSITE; PS50294; WD_REPEATS_REGION; 1.
KW Hypothetical protein; Repeat; WD repeat.
FT REPEAT 211..241 WD 1.
FT REPEAT 253..283 WD 2.
FT REPEAT 295..325 WD 3.
FT REPEAT 327..367 WD 4.
FT REPEAT 381..410 WD 5.
FT REPEAT 422..455 WD 6.
FT REPEAT 467..499 WD 7.
SQ SEQUENCE 501 AA; 54473 MW; DB22116996F85988 CRC64;

Query Match 14.0%; Score 361.5; DB 1; Length 501;
Best Local Similarity 24.8%; Pred. No. 8.3e-21;
Matches 123; Conservative 68; Mismatches 164; Indels 141; Gaps 17;

QY 11 TPHTTQARLYVTGEEAGPPIDLP--AGITTOQLGLICNALLKNEEATPYLFFVGEDEIK- 67
DB 116 TPHRVQSPVGL---LGPPGFPQFAGSTTSQ-----GNP-----SEVSVRT 154
QY 68 KSLEDTLD--LASVDENVIDIVYQPAQVFKVRPTVTRCTSS-----MP 108
DB 155 KSAEGTTGPIAPSTITTKPTSTIQVAPR---DPVAPTSSSGITKKPENGESFLVKTIS 210
QY 109 GHAEAVVSLNFSPDGAHLAGSGDVTVRDLNLTETPHCTCTGKQWVLCSWAPDGRKL 168
DB 211 GHKTSVSVKFSYCYGKYLGTGSADKQIKVNTVDMYTLQTLASHOLGINDFSWSSNSQFI 270
QY 169 ASGCKAGSIIIDWPEGQKGRPLSGHKKHINCLAWEPYHRDPCRLASASGDCGRTW 228
DB 271 ASASDVTTVKIFDVISGACL-RTRMGRHTNVVCCSF-----NPQSSLIASAGFETVRW 324
QY 229 DVKLGCLMNIAGHTNATVNRGAG--LIYSSKDRVTKMRAADGILCRTFSGHARW 287
DB 325 DEKTLGCLVKCI PAHSDPTITSINHDGNTWATSGCIRVWDASGCLTKL----- 377
QY 288 NNIALSTDYVLRGPHFPVDRSKSHLSLSTEEQLQESALKRYQAVCPDEVESLVSCSDN 347
DB 378 ---VDYDHA-----PV-----TFVC----- 389
QY 348 TLYLRNNQNKVERMTGHONVNDVKYSPDKLIASAFDKSYRLWRASDGYMATFRG 407
DB 390 -----FSPNGKYLISAQLDLSLKLWDPKAPKPLKYNG 422

```







GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 2, 2003, 06:17:40 ; Search time 59 Seconds  
(without alignments)  
1676.316 Million cell updates/sec

Title: US-09-830-980-1

Perfect score: 2574

Sequence: 1 MQETDTQETPTTIQARLV.....PDGSRVASGGKDKVILWAY 480

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_21:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2566	99.7	480	5	Q96995
2	2554.5	99.2	481	5	Q9VPR4
3	2554.5	99.2	488	5	Q8T4A2
4	1520	59.1	485	4	Q9NVX2
5	1520	59.1	487	4	Q9BU54
6	1513	58.8	476	13	Q93531
7	1512	58.7	485	11	Q8VEJ4
8	1335.5	51.9	473	10	Q9FLX9
9	1331	51.7	447	10	Q9AVI7
10	1318.5	51.2	473	5	Q9TVV3
11	1047.5	40.7	502	3	Q74855
12	616	23.9	1356	3	Q8X1P4
13	603	23.4	1356	3	Q8X1P5
14	603	23.4	1356	3	Q8X1P3
15	595	23.1	1376	3	Q8X1P2
16	538.5	20.9	1708	16	Q8YZI2

17	533	20.7	1227	16	Q8Z0R1	anabaena sp
18	532.5	20.7	1711	16	Q8Z0I9	anabaena sp
19	532	20.7	1747	16	Q8Z0Z0	anabaena sp
20	497	19.3	934	16	Q8YZ23	anabaena sp
21	496.5	19.3	1189	16	Q8YL09	anabaena sp
22	486	18.9	1551	16	Q8YMU3	anabaena sp
23	476.5	18.5	676	16	Q8YSG6	anabaena sp
24	469	18.2	1189	16	Q8YTD1	anabaena sp
25	462.5	18.0	1241	2	Q9XBD8	amycollatops
26	451	17.5	677	16	Q8YZL9	anabaena sp
27	436	16.9	304	16	Q8Z0S4	anabaena sp
28	425.5	16.5	1049	2	Q8ZEM4	streptomyc
29	425.5	16.5	1676	16	Q9KXX9	streptomyc
30	424.5	16.5	559	16	Q8YSC0	anabaena sp
31	412.5	16.0	700	5	Q9XZK1	drosophila
32	411	16.0	353	5	Q8T776	branchiosto
33	410.5	15.9	317	10	Q9M2Z2	arabidopsis
34	400	15.5	589	16	Q8YNI4	anabaena sp
35	399.5	15.5	414	5	Q8SW59	encephalito
36	399	15.5	330	4	Q9NUL4	homo sapien
37	396	15.4	613	10	Q9FNI9	arabidopsis
38	389	15.1	411	5	Q86698	drosophila
39	386.5	15.0	1051	17	Q8TMS3	methanosarc
40	385.5	15.0	328	11	Q9D7H2	mus musculu
41	377	14.6	522	4	Q9BQ87	homo sapien
42	372.5	14.5	514	4	Q9BZK7	homo sapien
43	372.5	14.5	514	4	Q9H9A1	homo sapien
44	371	14.4	786	16	Q8YNN6	anabaena sp
45	370.5	14.4	514	4	Q9H2I1	homo sapien

## ALIGNMENTS

RESULT 1

Q96995 ID Q96995 PRELIMINARY; PRT: 480 AA.

AC Q96995; DT 01-MAY-1999 (TREMBLrel. 10, Created)

DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)

DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)

DE NOTCHLESS protein.

GN NLE OR NOTCHLESS OR CG2863.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI\_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=99077802; PubMed=9857191;

RA Royet J., Bouwmeester T., Cohen S.M.;

RT "Notchless encodes a novel WD40-repeat-containing protein that

modulates Notch signaling activity.";

RL EMBO J. 17:7351-7360(1998).

CC -!- SIMILARITY: CONTAINS 8 WD REPEATS (TRP-ASP DOMAINS).

DR EMBL; AJ012588; CAA10070.1; -

DR FlyBase; FBgn0021874; Nle.

DR InterPro; IPR001632; Gprotein\_B.

DR Pfam; PF00400; WD40.

DR PRINTS; PR00319; GPROTEINB.

DR PRODOM; PD000018; WD40; 6.

DR SMART; SM00320; WD40; 8.

DR PROSITE; PS00678; WD\_REPEATS\_1; UNKNOWN\_2.

DR PROSITE; PS50082; WD\_REPEATS\_2; 7.

DR PROSITE; PS50294; WD\_REPEATS\_REGION; 1.

KW Repeat; WD repeat.

SQ SEQUENCE 480 AA; 52933 MW; 40A5D69D33956C4 CRC64;

Query Match 99.7%; Score 2566; DB 5; Length 480;  
Best Local Similarity 99.8%; Pred. No. 4.5e-212;

Matches 479; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MQETDTQEATPHITQARLVYTGEEAGPPIDLPAGITTOQLGICNALLKNEATPYLFF 60  
 DB 1 MQETDTQEATPHITQARLVYTGEEAGPPIDLPAGITTOQLGICNALLKNEATPYLFF 60

QY 61 VGEDEIKKSLEDTLDLASVDNTENVIVYQPAVFKVRPVRTCTSSMPGHAENVSLNFS 120  
 DB 61 VGEDEIKKSLEDTLDLASVDNTENVIVYQPAVFKVRPVRTCTSSMPGHAENVSLNFS 120

QY 121 PDGAHLASGGSDTIVRLWDLNTETPHFTCTGKHQWVLCVSWAPDGRKLASGCKAGSIILW 180  
 DB 121 PDGAHLASGGSDTIVRLWDLNTETPHFTCTGKHQWVLCVSWAPDGRKLASGCKAGSIILW 180

QY 181 DPFTGQKGRPLSGHKKHINCLAWPEYHRDPCEKRLASASGDDGDCRWDVKGQCLMNIA 240  
 DB 181 DPFTGQKGRPLSGHKKHINCLAWPEYHRDPCEKRLASASGDDGDCRWDVKGQCLMNIA 240

QY 241 GHTNAVTAVRWGAGLIYTSKDRVTWKWRAADGILCTFSGHAHVNNIALSTDYVLR 300  
 DB 241 GHTNAVTAVRWGAGLIYTSKDRVTWKWRAADGILCTFSGHAHVNNIALSTDYVLR 300

QY 301 GFPHVPKDRSKSHLSLSTEELESALKRYQAVCPDEVESLVSCDDNTLYLWRNNQKCV 360  
 DB 301 GFPHVPKDRSKSHLSLSTEELESALKRYQAVCPDEVESLVSCDDNTLYLWRNNQKCV 360

QY 361 ERMTGHQNVNDVKYSPDVKLIASAFDKSVRLWRASDGOYMATFRGHVQAVYTVANSAD 420  
 DB 361 ERMTGHQNVNDVKYSPDVKLIASAFDKSVRLWRASDGOYMATFRGHVQAVYTVANSAD 420

QY 421 SRLIVSGSKDSTLKVNSVQTKKLAQELPGHADEVFGVDWAPDGSRVASGCKDKVTKLWAY 480  
 DB 421 SRLIVSGSKDSTLKVNSVQTKKLAQELPGHADEVFGVDWAPDGSRVASGCKDKVTKLWAY 480

## RESULT 2

Q9VPR4 ID Q9VPR4 PRELIMINARY; PRT; 481 AA.

AC Q9VPR4;

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)

DE NLE protein.

GN NLE OR CG2863.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI\_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BERKELEY;

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Fandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blazee R.G., Champe M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA Abrell J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,

RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,

RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K.J., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,

RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,

RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,

RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Morkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,

RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA Spivak R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weltsenbach J.,

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Rao Q.A.,

RA Ye J., Yeh R.-F., Zaveril J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

RT "The genome sequence of Drosophila melanogaster.";

RL Science 287:2185-2195(2000).

CC -I- SIMILARITY: CONTAINS 8 WD REPEATS (TRP-ASP DOMAINS).

DR EMBL; AEO03589; RAF51479.1; .

DR Flybase; Fgn0021874; Nle.

DR InterPro; IPR001632; GproteIn\_B.

DR InterPro; IPR001680; WD40.

DR Pfam; PF00400; WD40; 9.

DR PRINTS; PR00319; GPROTEINB.

DR PRINTS; PR00320; GPROTEINBRPT.

DR ProDom; PD000018; WD40; 6.

DR SMART; SM00320; WD40; 8.

DR PROSITE; PS00678; WD\_REPEATS\_1; 2.

DR PROSITE; PS00682; WD\_REPEATS\_2; 7.

DR PROSITE; PS00294; WD\_REPEATS\_REGION; 1.

KW Repeat; WD repeat.

SQ SEQUENCE 481 AA; 52981 MW; F0EDD0815E691914 CRC64;

Query Match 99.2%; Score 2554.5; DB 5; Length 481;

Best Local Similarity 99.6%; Pred. No. 4.4e-211;

Matches 479; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 MQETDTQEATPHITQARLVY-TGEEAGPPIDLPAGITTOQLGICNALLKNEATPYLFF 59  
 DB 1 MQETDTQEATPHITQARLVSDTGEAGPPIDLPAGITTOQLGICNALLKNEATPYLFF 60

QY 60 FYGEDEIKKSLEDTLDLASVDNTENVIVYQPAVFKVRPVRTCTSSMPGHAENVSLNF 119  
 DB 61 FYGEDEIKKSLEDTLDLASVDNTENVIVYQPAVFKVRPVRTCTSSMPGHAENVSLNF 120

QY 120 SPGAHLASGGSDTIVRLWDLNTETPHFTCTGKHQWVLCVSWAPDGRKLASGCKAGSIIL 179  
 DB 121 SPGAHLASGGSDTIVRLWDLNTETPHFTCTGKHQWVLCVSWAPDGRKLASGCKAGSIIL 180

QY 180 WDPETGQKGRPLSGHKKHINCLAWPEYHRDPCEKRLASASGDDGDCRWDVKGQCLMNI 239  
 DB 181 WDPETGQKGRPLSGHKKHINCLAWPEYHRDPCEKRLASASGDDGDCRWDVKGQCLMNI 240

QY 240 AGHTNAVTAVRWGAGLIYTSKDRVTWKWRAADGILCTFSGHAHVNNIALSTDYVLR 299  
 DB 241 AGHTNAVTAVRWGAGLIYTSKDRVTWKWRAADGILCTFSGHAHVNNIALSTDYVLR 300

QY 300 TGPFPKDRSKSHLSLSTEELESALKRYQAVCPDEVESLVSCDDNTLYLWRNNQKNC 359  
 DB 301 TGPFPKDRSKSHLSLSTEELESALKRYQAVCPDEVESLVSCDDNTLYLWRNNQKNC 360

QY 360 VERMTGHQNVNDVKYSPDVKLIASAFDKSVRLWRASDGOYMATFRGHVQAVYTVANSA 419  
 DB 361 VERMTGHQNVNDVKYSPDVKLIASAFDKSVRLWRASDGOYMATFRGHVQAVYTVANSA 420

QY 420 DSRILVSGSKDSTLKVNSVQTKKLAQELPGHADEVFGVDWAPDGSRVASGCKDKVTKLWA 479  
 DB 421 DSRILVSGSKDSTLKVNSVQTKKLAQELPGHADEVFGVDWAPDGSRVASGCKDKVTKLWA 480

QY 480 Y 480  
 DB 481 Y 481

```
RESULT 3
Q874A2 PRELIMINARY; PRT; 488 AA.
ID Q874A2
AC Q874A2;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE AT08344p.
GN NLE.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champagne M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Celisier S.
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY089286; AAL90024.1; -.
SQ SEQUENCE 488 AA; 53809 MW; D4883202954A7785 CRC64;

Query Match
Best Local Similarity 99.2%; Score 2554.5; DB 5; Length 488;
Matches 479; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 MQETDTEQATPHITQARLVY--TGEAGPPIDLPAGITTOQLGLICNALLKNEEATPYLF 59
DB 8 MQETDTEQATPHITQARLVYSDTGEAGPPIDLPAGITTOQLGLICNALLKNEEATPYLF 67
QY 60 FYGEDETKKSLDLDLASVDTENVIDIYQAVFKVPVTRCTSSMPGHAENVLSNF 119
DB 68 FYGEDETKKSLDLDLASVDTENVIDIYQAVFKVPVTRCTSSMPGHAENVLSNF 127
QY 120 SPDGAAHLAGSGDPTVRLWDLNTEPHFTCTGHKQWVLCVSWAPDGKRLASGCKAGSIII 179
DB 128 SPDGAAHLAGSGDPTVRLWDLNTEPHFTCTGHKQWVLCVSWAPDGKRLASGCKAGSIII 187
QY 180 WDPETGQKGRPLSGHKKHINCLAWPEYHRDPECKRLASAGDGDGRINDVKLGOCIMNI 239
DB 188 WDPETGQKGRPLSGHKKHINCLAWPEYHRDPECKRLASAGDGDGRINDVKLGOCIMNI 247
QY 240 AGHTNAVTAVRWGAGLIYTSKDRITVKMRAADGILCTFSGHAHWNNIALSTDYVLR 299
DB 248 AGHTNAVTAVRWGAGLIYTSKDRITVKMRAADGILCTFSGHAHWNNIALSTDYVLR 307
QY 300 TGFPHFVDRSKSHLSLSTEELQESALKYQAVCPDEVESLVSGDDNTLYLWRNNQNC 359
DB 308 TGFPHFVDRSKSHLSLSTEELQESALKYQAVCPDEVESLVSGDDNTLYLWRNNQNC 367
QY 360 VERMTGCHQNVYNDVKYSPDVKLIASAFDVKSVRLWRASDGOYMATFRGHVQAVYVWASA 419
DB 368 VERMTGCHQNVYNDVKYSPDVKLIASAFDVKSVRLWRASDGOYMATFRGHVQAVYVWASA 427
QY 420 DSKLVSGSKDSTLKVMSVOTKKLAQLPGHADEVFGVDWAPDGSRSVAGKDKVILKWA 479
DB 428 DSKLVSGSKDSTLKVMSVOTKKLAQLPGHADEVFGVDWAPDGSRSVAGKDKVILKWA 487
QY 480 Y 480
DB 488 Y 488

RESULT 4
Q9NVX2 PRELIMINARY; PRT; 485 AA.
ID Q9NVX2
AC Q9NVX2;
```

```
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE CDNA FLJ10458 fis, clone NT2RP1001457, highly similar to Homo sapiens
DE partial beta-transducin family protein (Hypothetical 53.3 kDa
DE protein).
DE OS Homo sapiens (Human).
DE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
DE Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H., Sugawara M.,
RA Wagaatsuma M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA Ninomiya K., Iwavanagi T.;
RA "NEDO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Tissue-Placenta;
RA Strausberg R.;
RC Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
RL -1- SIMILARITY: CONTAINS 8 WD REPEATS (TRP-ASP DOMAINS).
DR EMBL; AK013205; BAA91621.1; -.
DR EMBL; BC012075; AAH12075.1; -.
DR InterPro; IPR001632; Gprotein_B.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 8.
DR PRINTS; PR00319; GPROTEINB.
DR PRINTS; PR00320; GPROTEINBRPT.
DR ProDom; PD000018; WD40; 7.
DR SMART; SM00320; WD40; 8.
DR PROSITE; PS00678; WD_REPEATS_1; UNKNOWN_4.
DR PROSITE; PS00682; WD_REPEATS_2; 7.
DR PROSITE; PS00294; WD_REPEATS_REGION; 1.
KW Hypothetical protein; Repeat; WD repeat.
SQ SEQUENCE 485 AA; 53266 MW; 0AEBE24B44957379 CRC64;

Query Match
Best Local Similarity 59.1%; Score 1520; DB 4; Length 485;
Matches 279; Conservative 73; Mismatches 118; Indels 6; Gaps 3;

QY 9 EATPHITQARLVYTGEEA---GPPIDLPAGITTOQLGLICNALLKNEEATPYLFFVGED 64
DB 8 EAVARDVQRLLYQFQDEGGQLLGSFDPVVDIIPDRLLQVNCALLAQEDPLPAFFVHDA 67
QY 65 EIKKSLDLDLASVDTENVIDIYQAVFKVPVTRCTSSMPGHAENVLSNFPDGA 124
DB 68 EIVSSLGKLTESQAVETEKVLDIIYQPAIFRVRVTRCTSSLEGHSEAVISVAFPTGK 127
QY 125 HLAGSGDPTVRLWDLNTEPHFTCTGHKQWVLCVSWAPDGKRLASGCKAGSIIIPDET 184
DB 128 YLAGSGDPTVRLWDLNTEPHFTCTGHKQWVLCVSWAPDGKRLASGCKAGSIIIPDET 187
QY 195 GQKGRPLSGHKKHINCLAWPEYHRDPECKRLASAGDGDGRINDVKLGOCIMNTAGHTN 244
DB 188 GKQGRPLSGHKKHINCLAWPEYHRDPECKRLASAGDGDGRINDVKLGOCIMNTAGHTN 247
QY 245 AVTAVRWGAGLIYTSKDRITVKMRAADGILCTFSGHAHWNNIALSTDYVLTGTPFH 304
DB 248 SVTCLRWGGLLYSASQDRTIKVRAHDGVLCRTLQGHGHWNTMALSTDYVLTGTFGE 307
QY 305 PVK--DRSKSHLSLSTEELQESALKYQAVCPDEVESLVSGDDNTLYLWRNNQ--KCVF 362
DB 308 PAEASVNPQDLQGLQELKERALSRLNVRGQGERLVSGDDFTLFLMSPAEDKKPLTR 367
QY 363 MTGCHQNVYNDVKYSPDVKLIASAFDVKSVRLWRASDGOYMATFRGHVQAVYVWASDR 422
DB 368 MTGCHQNVYNDVKYSPDVKLIASAFDVKSVRLWRASDGOYMATFRGHVQAVYVWASDR 427
```









```
QY 374 KYSDFVLIASAFDKSVLRASDGOYMATFRGHVQAVYTVANSADSLIVSGSKDSTL 433
Db 367 VSPDFRYIASAFDKSVKLCWGRGTGYLASRGRHGVGYVQVANSADSLIVSGSADSTL 426
QY 434 KWSVQTKLAQELPGHADEFGVDWAPDGRSVASGGDKVYKILW 478
Db 427 KVELKTSLYLDLPGHGVDFVTWNSPECTKVYSGGDKVYKILW 471

RESULT 11
O74855
ID O74855 PRELIMINARY; PRT; 502 AA.
AC O74855;
DT 01-NOV-1998 (TREMBLrel. 08, Created).
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Notchless-like, WD repeat protein.
GN SPC18.05C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN NCBI_TaxID=4896;
RP SEQUENCE FROM N.A.
RC STRAIN=972H-;
RA Hilbert H., Duesterhoeft A., Wood V., Rajandream M.A., Barrell B.G.;
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 9 WD REPEATS (TRP-ASP DOMAINS).
DR EMBL; AL031907; CAAT21419.1; -.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 8.
DR PRINTS; PR00320; GPROTEINBRPT.
DR ProDom; PD000018; WD40; 5.
DR SMART; SM00320; WD40; 8.
DR PROSITE; PS00678; WD_REPEATS_1; UNKNOWN_5.
DR PROSITE; PS0082; WD_REPEATS_2; 7.
DR PROSITE; PS0294; WD_REPEATS_REGION; 1.
KW Repeat.
SQ SEQUENCE 502 AA; 56003 MW; 5773208C7DEB23A CRC64;
```

```
Query Match 40.7%; Score 1047.5; DB 3; Length 502;
Best Local Similarity 45.3%; Pred. No. 1.9e-81;
Matches 208; Conservative 76; Mismatches 160; Indels 15; Gaps 6;

QY 32 LPAGITTOQLGLICNALKN-EEATPYLFVGEDEIKKSLEDTDLA-----SVDTENVI 85
Db 49 VPGNSVRLQALLNQLLENSDDVPYNFALHDEDETEIQDNLVTSVFHNGLMKTEDHL 108
QY 86 DIVYQFAVFRVPTVTRCTSSMPGHAEEAVVSLNFSFD- GAHLASGSDTTRVLDLNTET 144
Db 109 TLLYTPQAVFRVRAVTRCTASMGHDGTIIISAQFSPTSSRLVYSGDFTARLWCDTQT 168
QY 145 PHFTCTGKHQVLCVSWAPDGKRLASGCKAGSIIIWDPETGQOGRPLSGHKHINCLAW 204
Db 169 PIATMKHTNWVSCVAPADASIATGSDMNTIRFWDPKGSPIGDALRRHTKPIALCW 228
QY 205 EYHRDPECRK--LASASGDGDCRIWDVKLGQCLMNIAGHTNAVAVRMGGAGLIYTSK 262
Db 229 QPLHLAPDGGPYLLASGSKDNTVRIWVKLRLLFTLLSGHTAPITCVRWGQNNWYSSY 288
QY 263 DRTVKWRAADGILCRFTFSGHAHWNNIALSTDYVLRTPPHVPKRSKSHLSLSTELQ 322
Db 289 DKTIRWDKDGKCLHTLKGAARVNNHLSLSTEHVLRSGAYDHTDFKPKSF-----SDER 343
QY 323 ESALARYQAVCPDEVESLVCSDNTLYLW-RNNONKCVERTMCHONVNDVKYSPDKVL 381
Db 344 RKAKERYEACLKQSGRLVSASDDLQILWDPQKSTKPIIKMHGCHKVYNHASFDPGRC 403
QY 382 IASAFSDKSVLRASDGOYMATFRGHVQAVYTVANSADSLIVSGSKDSTLKVWSVQTK 441
Db 404 IATASFDSVRLWDGRTGKFLATLRGHVAAYVQCAWSTDSRLVSSQDTTLKVDVRSK 463
```

```
QY 442 KLAQELPGHADEFGVDWAPDGRSVASGGDKVYKILW 480
Db 464 KMKFDLPGHEDQVFAVDWSPDQGVASGADKAVRIWSH 502

RESULT 12
O8X1P4
ID O8X1P4 PRELIMINARY; PRT; 1356 AA.
AC O8X1P4;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Beta transducin-like protein HET-E2C*4.
GN HET-E.
OS Podospora anserina.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariatales; Lasiosphaeriaceae; Podospora.
OX NCBI_TaxID=5145;
RN NCBI_TaxID=5145;
RP SEQUENCE FROM N.A.
RA Espagne E., Balhadere P., Bequeret J., Turcq B.;
RT "A new family of WD40 proteins implicated in vegetative incompatibility: evidence for a major role of WD40 repeat domain in the specificity of het-e and het-d genes.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF323583; AAL37299.1; -.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 10.
DR PRINTS; PR00320; GPROTEINBRPT.
DR ProDom; PD000018; WD40; 10.
DR SMART; SM00320; WD40; 10.
DR PROSITE; PS00678; WD_REPEATS_1; UNKNOWN_10.
DR PROSITE; PS0082; WD_REPEATS_2; 10.
DR PROSITE; PS0294; WD_REPEATS_REGION; 1.
SQ SEQUENCE 1356 AA; 149702 MW; 33148AF47B82826 CRC64;

Query Match 23.9%; Score 616; DB 3; Length 1356;
Best Local Similarity 29.4%; Pred. No. 1.1e-43;
Matches 150; Conservative 86; Mismatches 203; Indels 72; Gaps 10;

QY 13 HTIQARLVYTGEEAGPPIDLPAGITTOQLGLICNALKNKEATPY---LFFVGEDEIKKS 69
Db 763 HTINGRLIALVRDG-----YFALSVMKIEKAPQAVTSALVAPDTSMIKK 810
QY 70 LEDFLDLASVDTEENVIDIVYQFAVFRVPTVTRCTSSMPGHAEEAVVSLNFSFDGAHLASG 129
Db 811 IFKKEEPCWISTISVVEAEW-----NACTQTLEHGSGSVLSVAFSPDQGVASG 859
QY 130 SGDTTVRLDLNLTETPHFTCTGKHQVLCVSWAPDGKRLASGCKAGSIIIWDPETGQOGRK 189
Db 860 SDDTKIKWDTATSGTGTQTLLEGHGSVWSVAFSPDQGVASGDDKTIKIMDAASGCT- 918
QY 190 RPLSGHKKHINCLANEPYHR-----DPE 212
Db 919 QTLLEGHGSVLSVAFSPDQGVASGDKTIKIMDTASGTGTQTLLEGHGSVWSVAFSPD 978
QY 213 CRKLASASGDGDCRIWDVKLGQCLMNIAGHTNAVAVRMGGAG-LIYTSKDRVTKWRA 271
Db 979 GQVVASGSGDKTIKIMDTASGTGTQTLLEGHGSVWSVAFSPDQGVASGDDKTIKIMDT 1038
QY 272 ADGILCRFTFSGHAHWNNIALSTD--YVLRTPPHVPK--DRSKSHLSLSTELQESALK 327
Db 1039 ASGTQTQTLLEGHGWYQSVAFSPDQGVASGNDHTIKIMDAASGTCTQTLLEGHGSVWS 1098
QY 328 RYQAVCPDEVESLVCSDNTLYLWRNNONKCVERTMCHONVNDVKYSPDKVLASASF 387
Db 1099 --VAFSPDG-QRVASGDDHTIKIMDAASGTCTQTLLEGHGSVWSVAFSPDQGVASGI 1155
QY 388 DKSRLWRASDGOYMATFRGHVQAVYTVANSADSLIVSGSKDSTLKVWSVQTKLAQEL 447
Db 1156 DGTIKIMDAASGTCTQTLLEGHGWVHVSFAVSPDQGVASGSDGTIKIMDAASGTCTOTL 1215
QY 448 PGHADEFGVDWAPDGRSVASGGDKVYKILW 478
```



## RESULT 15

```
Q8X1P2 PRELIMINARY; PRT; 1376 AA.
ID Q8X1P2;
AC Q8X1P2;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Beta transducin-like protein HET-D2Y.
GN HET-D.
OS Podospora anserina.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Lasiosphaeriaceae; Podospora.
OX NCBI_TaxID=5145;
RN [1]
RP SEQUENCE FROM N.A.
RA Espagne E., Balhadere P., Begueret J., Turcq B.;
RT "A new family of WD40 proteins implicated in vegetative
RT incompatibility; evidence for a major role of WD40 repeat domain in
RT the specificity of het-e and het-d genes.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF323585; AAL37301.1; -
DR InterPro: IPR001680; WD40.
DR Pfam: PF00400; WD40; 12.
DR PRINTS: PR00320; GPROTEINRPT.
DR PRODOM: PD000018; WD40; 9.
DR SMART: SM00320; WD40; 12.
DR PROSITE: PS00678; WD_REPEATS_1; UNKNOWN_11.
DR PROSITE: PS00082; WD_REPEATS_2; 11.
DR PROSITE: PS0294; WD_REPEATS_REGION; 1.
SQ SEQUENCE 1376 AA; 152068 MW; 2191826AA5C62D97 CRC64;

Query Match 23.1%; Score 595; DB 3; Length 1376;
Best Local Similarity 34.1%; Pred. No. 7, le-42;
Matches 131; Conservative 66; Mismatches 167; Indels 20; Gaps 6;

QY 103 CTSMPGHAENAVSLNFPDGAHLASGSDTVRLDLNLTETPHFTCTCHQWVLCVSWA 162
DB || || || || || || || || || || || || || || || || || || || || ||
QY 865 CTQTLEHGHWLVSFAFSPDSKWWASGSADSTIKIWEAATGCTQTLEHGHWVSVAFS 924
DB || || || || || || || || || || || || || || || || || || || || ||
QY 163 PDGRKLASGCKAGSIIIDPETQQQGRPLSGCHKHKNCLANPEYHRDPECKKLASASGD 222
DB || || || || || || || || || || || || || || || || || || || || ||
QY 925 PDSKWWVSGSADSTIKIWEAATGCTQTLEHGHWVSVAF-----SPDSKWWASGSAD 978
DB || || || || || || || || || || || || || || || || || || || || ||
QY 223 GDCRIWDVKLGQCLMNIAGHTNAVAV-----PWGAGLIYTSSKDRIVKWRADGIL 276
DB || || || || || || || || || || || || || || || || || || || || ||
QY 979 STIKIWEAATGCTQTLEHGHWVSVAFSPDSKWWASG-----SDHTIKIWEAATGSC 1033
DB || || || || || || || || || || || || || || || || || || || || ||
QY 277 CRTFSGHAWYNNIALSTD--YVLRTPFPVKDRSKSHLSLSTELQESALKRYQAVCP 334
DB || || || || || || || || || || || || || || || || || || || || ||
QY 1034 TQTLEHGHWVSVAFSPDSKWWASGSDDHTIKIWEAATGS-CTQTLEHGHWVSVAFS 1092
DB || || || || || || || || || || || || || || || || || || || || ||
QY 335 DVEISLVSCDDNTLYLRNNONKCVERTGHONVNDVYSPDKVLIASAFDKSVRLW 394
DB || || || || || || || || || || || || || || || || || || || || ||
QY 1093 PDSKWWVSGSADSTIKIWEAATGCTQTLEHGHWVSVAFSPDSKWWASGSTDTIKIW 1152
DB || || || || || || || || || || || || || || || || || || || || ||
QY 395 RASDQGYWATPRGHVQAVYTVAMSADSLIVSGSKDSTLKWVSVQTKLAQLPGHADEV 454
DB || || || || || || || || || || || || || || || || || || || || ||
QY 1153 EAATGCTQTLEHGHWVSVAFSPDSKWWASGSADSTIKIWEAATGCTQTLEHGHWGPV 1212
DB || || || || || || || || || || || || || || || || || || || || ||
QY 455 FGVDWAPDGSRVASGDKVKILW 478
DB || || || || || || || || || || || || || || || || || || || || ||
QY 1213 NSVAFSPDSKWWASGSDDHTIKIW 1236
DB || || || || || || || || || || || || || || || || || || || || ||
```

Search completed: June 2, 2003, 07:00:48  
Job time : 63 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 2, 2003, 03:21:05 ; Search time 3984 Seconds  
(without alignments)  
11359.152 Million cell updates/sec

```

Title: US-09-830-980-2
Perfect score: 1555
Sequence: 1 aattcccaaaaaatgcagga.....aaaaaaaaaaaaaaaaaaaaa 1555

```

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
C	1	1555	100.0	1555	6	AX025192	Sequence
	2	1528.2	98.3	1663	3	AY089286	Drosophil
	3	1521	97.8	1521	3	DM012588	Drosophil
	4	1285.8	82.7	85095	3	AC004573	Drosophil
	5	1285.8	82.7	142257	2	AC0017491	Drosophil
C	6	1285.8	82.7	182726	3	AC008002	Drosophil
	7	1285.8	82.7	302473	3	AE003589	Drosophil
	8	478	30.7	1859	6	AC135811	Sequence
	9	477.4	30.7	1853	9	AK001320	Homo sapi
	10	477.4	30.7	1854	6	AX077627	Sequence
C	11	477.4	30.7	1870	9	BC002884	Homo sapi
	12	477.4	30.7	2593	9	BC012075	Homo sapi
	13	477.4	30.7	2638	6	AX045715	Sequence
	14	476.6	30.6	1784	10	BC018399	Mus muscu
	15	466.2	30.0	1928	6	AF069737	Sequence
C	16	464.6	29.9	2250	5	AF069737	Xenopus l
	17	249.4	16.0	1170	6	AX135809	Sequence
	18	210.8	13.6	316613	8	SCCHR111	X59720. S. cerevisia
	19	159.8	10.3	1563	6	AX488920	Sequence
	20	120.4	7.7	378	6	AX072820	Sequence
C	21	115.8	7.4	40305	8	SPCC18	Sequence
	22	111.6	7.2	1373	9	HSNJ5257	S.pombe c
	23	90.4	5.8	31787	3	AF106576	Homo sapi
	24	87.8	5.6	215786	9	AC022916	Homo sapi
	25	86.2	5.5	203769	2	AL603745	Mus muscu
C	26	86.2	5.5	216655	2	AC120837	Mus muscu
	27	84.6	5.4	132348	10	CNS07Y02	AL713882 Mus muscu
	28	82.8	5.3	334520	1	AP003588	Nostoc sp
	29	79.8	5.1	34131	2	AC096796	Rattus no
	30	79.8	5.1	178633	2	AC115671	Rattus no
C	31	79.4	5.1	1844	9	BC025080	Homo sapi
	32	79.4	5.1	333550	1	AP003593	Nostoc sp
	33	78	5.0	4722	8	AF132585	Podospora
	34	77	5.0	4619	1	AF115313	Thermomon
	35	76.2	4.9	2740	9	HSM042229	AB009055 Arabidops
C	36	76.2	4.9	81575	8	AB009055	AB009055 Arabidops
	37	74.4	4.8	2311	3	AY113628	Homo sapi
	38	73.6	4.7	5413	8	AF132583	Drosophill
	39	72.8	4.7	1209	3	AF100287	Trypanoso
	40	72	4.6	5413	8	AF132582	Podospora
C	41	72	4.6	5413	8	AF132584	Podospora
	42	72	4.6	10280	1	AE012949	Chlorobriu
	43	71	4.6	600	11	CNS06EEP	L38125 Podospora a
	44	70.6	4.5	5411	8	FLA81PLP	L38125 Podospora a
	45	68.4	4.4	2977	6	AX430450	Sequence

## ALIGNMENTS

RESULT	1
LOCUS	AX025192
DEFINITION	Drosophila sp.
ACCESSION	Sequence 9 from Patent WO0026364.
VERSION	AX025192
KEYWORDS	AX025192.1 GI:10186913
SOURCE	.
ORGANISM	Drosophila sp. Drosophila sp. Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Epiphytoidea; Drosophilidae; Drosophila. 1 (bases 1 to 155)
REFERENCE	Cohen, S., Bouwmeester, A. and Royet, J. Regulator of notch signaling activity
AUTHORS	
TITLE	

JOURNAL Patent: WO 0026364-A 9 11-MAY-2000;  
EUROP MOLECULAR BIOLOGY LAB (DE) ; COHEN STEPHEN (DB) ; BOUWMEESTER  
ANTONIUS (DE) ; ROYET JULIEN (FR)

FEATURES Location/Qualifiers

source 1..1555  
/organism="Drosophila sp."  
/db\_xref="taxon:7242"

BASE COUNT 418 a 373 c 448 g 316 t

ORIGIN

Query Match 100.0%; Score 1555; DB 6; Length 1555;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1555; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AATTCCCAAAAAATCGACGAGACGGACAGAGGCCACGCACATCAGTACAG 60  
| | | | |  
Db 1 AATTCCCAAAAAATCGACGAGACGGACAGAGGCCACGCACATCAGTACAG 60  
| | | | |

QY 61 CGCGCTCGTTTACACGGCGAGAAAGCGGCCCAATCGACTCGCGGACGAATC 120  
| | | | |  
Db 61 CGCGCTCGTTTACACGGCGAGAAAGCGGCCCAATCGACTCGCGGACGAATC 120  
| | | | |

QY 121 ACTACCAGCAATTTGGGACTGATTTGCAACGCGCTGTGAAAACAGAGAACCTCCA 180  
| | | | |  
Db 121 ACTACCAGCAATTTGGGACTGATTTGCAACGCGCTGTGAAAACAGAGAACCTCCA 180  
| | | | |

QY 181 TATTGTTTTTCGTGGCGGAGTAGATCAAGAAGACCTGGAGGACACGTTGGACTTG 240  
| | | | |  
Db 181 TATTGTTTTTCGTGGCGGAGTAGATCAAGAAGACCTGGAGGACACGTTGGACTTG 240  
| | | | |

QY 241 GCSTCAGTGGACACCGAAACGTCATCGATTGTTGATCACGCCACAGCGGTTTTCAA 300  
| | | | |  
Db 241 CGTTCAGTGGACACCGAAACGTCATCGATTGTTGATCACGCCACAGCGGTTTTCAA 300  
| | | | |

QY 301 GTGCGCCAGTGACAAGATGCACAGATTCCATCCGCGGACAGCGGAGGCTGTGGTTTCG 360  
| | | | |  
Db 301 GTGCGCCAGTGACAAGATGCACAGATTCCATCCGCGGACAGCGGAGGCTGTGGTTTCG 360  
| | | | |

QY 361 CTGAATTTACGCCCGGATGGTGCATCTCCGCAGTGGAAAGTGCGCACACAGTGCGA 420  
| | | | |  
Db 361 CTGAATTTACGCCCGGATGGTGCATCTCCGCAGTGGAAAGTGCGCACACAGTGCGA 420  
| | | | |

QY 421 TTGTGGGATCTTAACACAGACACCGCACTTCACCTGCACAGGTTCATAAGCAGTGGGTT 480  
| | | | |  
Db 421 TTGTGGGATCTTAACACAGACACCGCACTTCACCTGCACAGGTTCATAAGCAGTGGGTT 480  
| | | | |

QY 481 CTGTGCGTATCTTGGCTCCGGATGGCAACGTTGGCCAGCGGTTGCCAAGCGGCTCT 540  
| | | | |  
Db 481 CTGTGCGTATCTTGGCTCCGGATGGCAACGTTGGCCAGCGGTTGCCAAGCGGCTCT 540  
| | | | |

QY 541 ATAATCATCTGGGACCCGGACAGCGGTCAGCAAGAGGGCGACCCCTTGAGTGGGCACAAG 600  
| | | | |  
Db 541 ATAATCATCTGGGACCCGGACAGCGGTCAGCAAGAGGGCGACCCCTTGAGTGGGCACAAG 600  
| | | | |

QY 601 AAACACATCAACTGCTCGCTCGGTAACCGCTATCATCGATCCGAGTGCAGGAACTT 660  
| | | | |  
Db 601 AAACACATCAACTGCTCGCTCGGTAACCGCTATCATCGATCCGAGTGCAGGAACTT 660  
| | | | |

QY 661 GCTTCCGCCAGTGGAGACGGGACTCGCCGATTGGGACGTAAAAATTGGGCCAGTGCCTT 720  
| | | | |  
Db 661 GCTTCCGCCAGTGGAGACGGGACTCGCCGATTGGGACGTAAAAATTGGGCCAGTGCCTT 720  
| | | | |

QY 721 ATGAACATTCGCGGACACACAATTCGTGACAGCAGTGCAGATGGGTTGGAGCGGCTT 780  
| | | | |  
Db 721 ATGAACATTCGCGGACACACAATTCGTGACAGCAGTGCAGATGGGTTGGAGCGGCTT 780  
| | | | |

QY 781 ATTATATACCTCTAAAGATCGCAGTGAAGATGTGGCAGCAGCTGTATGGNACTTGT 840  
| | | | |  
Db 781 ATTATATACCTCTAAAGATCGCAGTGAAGATGTGGCAGCAGCTGTATGGNACTTGT 840  
| | | | |

QY 841 TGCGGGAGCTTCTTGCGCGCAAGCTCACTGGGTAAACAACATTGCGCTCAGCACCGATTAC 900  
| | | | |  
Db 841 TGCGGGAGCTTCTTGCGCGCAAGCTCACTGGGTAAACAACATTGCGCTCAGCACCGATTAC 900  
| | | | |

QY	901	GTCTCGCGCAGTGGTCCATTCATCCGGTGAAGGATCGCTCCAGAGGACCACTCAGTTTG	960
Db	901		960
QY	961	AGCACTCAGGAATTCGAGGAATTCGCTTTGAAGCGCTACACAGGCCGTGTGCGCTGACGAG	1020
Db	961	AGCACTCAGGAATTCGAGGAATTCGCTTTGAAGCGCTACACAGGCCGTGTGCGCTGACGAG	1020
QY	1021	GTGGAGTCGCTGGTTTCCTGTTTCGGATGACAACACCCCTCTATCTGTGGCGGGAACAACAG	1080
Db	1021	GTGGAGTCGCTGGTTTCCTGTTTCGGATGACAACACCCCTCTATCTGTGGCGGGAACAACAG	1080
QY	1081	ACAAGTCGCTGGAGCGATGACAGGCGCACCAAGACGTGTCAACAGATGTGAATATTCG	1140
Db	1081	ACAAGTCGCTGGAGCGATGACAGGCGCACCAAGACGTGTCAACAGATGTGAATATTCG	1140
QY	1141	CCGATCTAAAGCTAATTCGCTGTCTTCATTTGACAAGTCAGTCGCTGTGTGGCGAGCC	1200
Db	1141	CCGATCTAAAGCTAATTCGCTGTCTTCATTTGACAAGTCAGTCGCTGTGTGGCGAGCC	1200
QY	1201	AGCATGGTCAGTACATGCGACCTCCGGGGTCACTGTGACAGGCTGTTTACACGGTGGCC	1260
Db	1201	AGCATGGTCAGTACATGCGACCTCCGGGGTCACTGTGACAGGCTGTTTACACGGTGGCC	1260
QY	1261	TGTCGCGGAGCTCCGCTTGATTTGTTTCGGCGCAAGACTCAACTCTAAAGTATGG	1320
Db	1261	TGTCGCGGAGCTCCGCTTGATTTGTTTCGGCGCAAGACTCAACTCTAAAGTATGG	1320
QY	1321	AGTGTGCAGACGAAGAAATCGCACAGGAGCTCCCTGGACATGCGGATGAGTGTTCGGA	1380
Db	1321	AGTGTGCAGACGAAGAAATCGCACAGGAGCTCCCTGGACATGCGGATGAGTGTTCGGA	1380
QY	1381	GTGACATGGCGCCCGATGGCTCTAGAGTTGCTCTGTGGTGGCAAGGACAAAGTTATAAG	1440
Db	1381	GTGACATGGCGCCCGATGGCTCTAGAGTTGCTCTGTGGTGGCAAGGACAAAGTTATAAG	1440
QY	1441	CTATGGGCTTATTAAACAAATCAATTAATCTGTACAGGTAAAGAAATACTTGAAGATAAG	1500
Db	1441	CTATGGGCTTATTAAACAAATCAATTAATCTGTACAGGTAAAGAAATACTTGAAGATAAG	1500
QY	1501	TAAACGCTCTGAGTAAAAA	1555
Db	1501	TAAACGCTCTGAGTAAAAA	1555
RESULT 2			
LOCUS	AY089286	1663 bp	linear
DEFINITION	Drosophila melanogaster	AY08344	full insert cDNA.
ACCESSION	AY089286		
VERSION	AY089286.1	GI:19527818	
KEYWORDS	FLI_CDNA.		
SOURCE	Drosophila melanogaster.		
ORGANISM	Drosophila melanogaster.		
REFERENCE	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephyroidea; Drosophilidae; Drosophila.		
AUTHORS	1 (bases 1 to 1663)		
TITLE	Stapleton, M., Brokstein, P., Hong, L., Agbayani, A., Carlson, J.,		
JOURNAL	Submitted (12-MAR-2002) Berkeley Drosophila Genome Project,		
	Lawrence Berkeley National Laboratory, One Cyclotron Road,		
	Berkeley, CA 94720, USA		
COMMENT	Sequence submitted by:		
	Lawrence Berkeley National Laboratory		
	Berkeley, CA 94720		



This clone was sequenced as part of a high-throughput process to sequence clones from Drosophila Gene Collection 1 (Rubin et al., Science 2000). The sequence has been subjected to integrity checks for sequence accuracy, presence of a polyA tail and contiguity within 100 kb in the genome. Thus we believe the sequence to reflect accurately this particular cDNA clone. However, there are artifacts associated with the generation of cDNA clones that may have not been detected in our initial analyses such as internal priming, priming from contaminating genomic DNA, retained introns due to reverse transcription of unspliced precursor RNAs, and reverse transcriptase errors that result in single base changes. For further information about this sequence, including its location and relationship to other sequences, please visit our Web site (<http://fruitfly.berkeley.edu>) or send email to [cdna@fruitfly.berkeley.edu](mailto:cdna@fruitfly.berkeley.edu).

## FEATURES

Location/Qualifiers

1..1663  
 /organism="Drosophila melanogaster"  
 /db\_xref="FLYBASE:FBgn0021874"  
 /map="21C6-21C5"

## gene

1..1663  
 /gene="Nle"

/note="alignment with genomic scaffold AB003589"

/db\_xref="FLYBASE:FBgn0021874"

100..1566

/gene="Nle"

/note="Longest ORF"

/codon\_start=1

/product="AT08344p"

/protein\_id="AAL50024.1"

/db\_xref="GI:19527819"

/db\_xref="FLYBASE:FBgn0021874"

/translation="MLAKKQMOETDEQEPHTIQRILVSDTGEAEPPIDLPAGI  
 TTQQLGLICNALLKNEETPYLFVEGDEIKKLEDTLDLASVDENVIDIVPOAV  
 FKVPRVTRCTSSMPGHAEEAVVSNFSPGHAHLAGSGDTTVRLWDLNTEPHETCTGH  
 KOWILCVSWPDGKRLASGCKAGSIIDWPETGQKGPRLSGHKHKLNCIAEPVHRD  
 PECKLASAGDGDRIWVKLGQCLMNIAGHTNAVAVRWGAGLIYSSDRVRKM  
 WRAADGILCRTEFSCHAWNNIALSTDYVLRGTFPHVVKDRSKHLSLSEELQESAL  
 KRYQAVCFDEVESLSCSDNTILWNRNQKNCVERFMGHONVNVNPKYSPDVKLIAS  
 ASFDKSLRWLRASDGTMTFRGHQVAVTYVAMSADSLIVSGSKDSTLKWVSVQTK  
 LAQELPHADEVFQVWAPDGRSVASGGKRVKILWAY"

BASE COUNT 444 a 397 c 478 g 344 t

## ORIGIN

Query Match 98.3%; Score 1528.2; DB 3; Length 1663;

Best Local Similarity 99.6%; Pred. No. 0;

Matches 1543; Conservative 0; Mismatches 3; Indels 3; Gaps 1;

QY 7 CAAAATCGAGGACGGACACGGAGCAAGAGCGCCCAATCGACCTGCCGCGAGGAATCACT 123

DB 115 CAGAAATCGAGGACGGACACGGAGCAAGAGCGCCCAATCGACCTGCCGCGAGGAATCACT 174

QY 67 CTGCTTT---ACACGCGGAGGAAGCGCGCCCAATCGACCTGCCGCGAGGAATCACT 123

DB 175 CTGCTTTCGGACACGGGAGGAGCGCGCCCAATCGACCTGCCGCGAGGAATCACT 234

QY 124 ACCGCAATTTGGGACTGATTTCGAACGGCTGCTGTAAGAAACGAGGAGCCATCCATAT 183

DB 235 ACCGCAATTTGGGACTGATTTCGAACGGCTGCTGTAAGAAACGAGGAGCCATCCATAT 294

QY 184 TTGTTTTTCGTCGGCGAGGATGAGATCAAGAGAGCGCTGGAGGACGCTTGGACTTGGCG 243

DB 243 TTGTTTTTCGTCGGCGAGGATGAGATCAAGAGAGCGCTGGAGGACGCTTGGACTTGGCG 354

QY 303 TTGGACCGGAAACGTTGATCGATATTTGTTATCAGCCACAGCGGCGTTTCAAAGTG 303

DB 414 TTGGACCGGAAACGTTGATCGATATTTGTTATCAGCCACAGCGGCGTTTCAAAGTG 414

QY 363 GATGCACGAGTTCCATGCGCGGACACCGCGAGGCTGTGGTTTCGCTG 363

DB 474 GATGCACGAGTTCCATGCGCGGACACCGCGAGGCTGTGGTTTCGCTG 474

CTCAGGAGTGGAGTGGCGACACCAACAGTGGCGATTG 423

DB 475 AATTTACGCGCGGATGGTGTCTATCTCGCCAGTGGAACTGGCGACACCAACAGTGGCGATTG 534

QY 424 TGGGATCTTAACACAGAGACACCGCCTTCCCTGCACAGGTCTATAAGCAGTGGGTTCTG 483

DB 535 TGGGATCTTAACACAGAGACACCGCCTTCCCTGCACAGGTCTATAAGCAGTGGGTTCTG 594

QY 484 TCGGTATCTCTGGGCTCGGATGCAACCGTTGGCCACGGTTGCCAAGCGGCTCTATA 543

DB 595 TCGGTATCTCTGGGCTCGGATGCAACCGTTGGCCACGGTTGCCAAGCGGCTCTATA 654

QY 544 ATCATCTGGGACCGGAGACGGGTACAGAGAGGGCGACCTTTGAGTGGCGACAAGAAA 603

DB 655 ATCATCTGGGACCGGAGACGGGTACAGAGAGGGCGACCTTTGAGTGGCGACAAGAAA 714

QY 604 CACATCAACTGCCTCGCTGGGAACCGTATCATCGGATCCGGAGTCCAGGAACCTTGTCT 663

DB 715 CACATCAACTGCCTCGCTGGGAACCGTATCATCGGATCCGGAGTCCAGGAACCTTGTCT 774

QY 664 TCCGCCAGTGGAGACGGGACTCCCGGATTTGGGACGTAAATTTGGGCCAGTGCCTTATG 723

DB 775 TCCGCCAGTGGAGACGGGACTCCCGGATTTGGGACGTAAATTTGGGCCAGTGCCTTATG 834

QY 724 AACATTGCGCGACACACAAATGCTGTGACAGCAGTGAGATGGGTGGCGAGCGGCTTATT 783

DB 835 AACATTGCGCGACACACAAATGCTGTGACAGCAGTGAGATGGGTGGCGAGCGGCTTATT 894

QY 784 TATACATCTCTCCAAAGATGCACAGTGAAGATGTGGGAGCAGCTGATGAAATCTTTGTGC 843

DB 895 TATACATCTCTCCAAAGATGCACAGTGAAGATGTGGGAGCAGCTGATGAAATCTTTGTGC 954

QY 844 CGGAGCTTCTCTGCCAAGCTCACTGGGTAAACAACATTTGGCTTGAGCAGCCAGTACGTC 903

DB 955 CGGAGCTTCTCTGCCAAGCTCACTGGGTAAACAACATTTGGCTTGAGCAGCCAGTACGTC 1014

QY 904 CTGCGCACTGGTCCATTCATCCATCCGTTGAAGATGCTCCAGAGCACCCTCATGTTGAGC 963

DB 1015 CTGCGCACTGGTCCATTCATCCATCCGTTGAAGATGCTCCAGAGCACCCTCATGTTGAGC 1074

QY 964 ACTGAGGAATTTGAGGAATCTGCTTGAAGCGCTACACAGCGGCTGTGCGCTGACGAGGTG 1023

DB 1075 ACTGAGGAATTTGAGGAATCTGCTTGAAGCGCTACACAGCGGCTGTGCGCTGACGAGGTG 1134

QY 1024 GAGTCGCTGTTTCTGTTGGATGACAAACACCTCTATCTGTGGCGGAGCAACACAGAAC 1083

DB 1135 GAGTCGCTGTTTCTGTTGGATGACAAACACCTCTATCTGTGGCGGAGCAACACAGAAC 1194

QY 1084 AAGTCGCTTGAGCGCATGACAGGCGCACAGAACCTGTGTCAACGATGTGAATATTTCCGCG 1143

DB 1195 AAGTCGCTTGAGCGCATGACAGGCGCACAGAACCTGTGTCAACGATGTGAATATTTCCGCG 1254

QY 1144 GATCTAAGCTAATTTGCTGCTGTTTCAATTTGCAAGTCAGTGGCTGTGTGGCGAGCCAGC 1203

DB 1255 GATCTAAGCTAATTTGCTGCTGTTTCAATTTGCAAGTCAGTGGCTGTGTGGCGAGCCAGC 1314

QY 1204 GATGTCAGTACATGCGCCACCTTCCGGGCTCATGTGAGGCTGTTTACAGCTGTCCCTGG 1263

DB 1315 GATGTCAGTACATGCGCCACCTTCCGGGCTCATGTGAGGCTGTTTACAGCTGTCCCTGG 1374

QY 1264 TCCGCGACCTCCGCTGTTGTTTCCGGCGACAGAACACTCAACTCTAAAGATATGGAGT 1323

DB 1375 TCCGCGACCTCCGCTGTTGTTTCCGGCGACAGAACACTCAACTCTAAAGATATGGAGT 1434

QY 1324 GTGAGACGAGAAACATGGCACAGGAGTGCCTGGACATCGGATCAGGTGTTCGGAGTG 1383

DB 1435 GTGAGACGAGAAACATGGCACAGGAGTGCCTGGACATCGGATCAGGTGTTCGGAGTG 1494

QY 1384 GACTGGGCGCGCATGGCTCTAGCTTCCCTCTGGTGGCGAGCAAGACTATATAAGCTA 1443

DB 1495 GACTGGGCGCGCATGGCTCTAGCTTCCCTCTGGTGGCGAGCAAGACTATATAAGCTA 1554

QY 1444 TGGGCTTATTAAACAAATCAATTAACTTTTACACGGTAAAGAAATACTTAGGAATAAAGTAA 1503

ACGGTAAGAAATACCTAGGAATAAGTAA 1614

AAAAAAAAAAAAAAAAAAAA 1552

AAAAAAAAAAAAAAAAAAAA 1663

1 bp mRNA linear INV 12-MAR-1999  
NA for Notchless protein.

ats.

.hropoda; Hexapoda; Insecta; Pterygota;  
a; Diptera; Brachycera; Muscomorpha;  
idae; Drosophila.

REFERENCE  
AUTHORS Royet J., Bouwmeester T. and Cohen S.M.  
TITLE Notchless encodes a novel WD40-repeat-containing protein that  
modulates Notch signaling activity  
JOURNAL EMBO J. 17 (24), 7351-7360 (1998)  
MEDLINE 99077802  
PUBMED 9857191

Location/Qualifiers  
1..1521  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"

1..1443  
/gene="notchless"

1..1443  
/gene="notchless"

1..1443  
/note="WD40 repeat containing protein"

1..1443  
/product="Notchless protein"

1..1443  
/protein\_id="CAAL0070.1"

1..1443  
/db\_xref="GI:4127781"

1..1443  
/translation="MOETDTEQATPHTIQARIVYVTEEGAGPPIDLPAGITQOGLI  
CNALKNEEATPYLPFWGEDEIKSLIEDLDSVDFENVIVPOAVERPVR  
CTSSMPGHAERYVSLNFPDGAHLAGSGDTIVRLDNIETPHEFTCGHKWVLCVS  
WAPDGKRIASGCKRAGSIIIDPETQQQRPLSGHKHINCLAWEPYHRDPCKRLVS  
ASGDGCRIDWVKGLQCLMNIAGHTNAVAVRWGAGLIYTSKDRYHMMRAADGIL  
CRTFSQAHVNNIALSTDYLVRTGPFPPVDRSKSLSTELQESALKRYQAVCP  
DEVSLVSCDDNTLYLWRNNQNKVERMTGHQNVNDVYKSPDKLIASAFKRSVR  
LWRASDGOYMATFRGHVQAVTVAMSDSRILVSGSKDSLKWSVQTKLAQELPGH  
ADEVFGVDMAPDGSRVASGDKVILKWAY"

BASE COUNT 389 a 370 c 448 g 314 t

ORIGIN

Query Match 97.8%; Score 1521; DB 3; Length 1521;  
Best Local Similarity 100.0%; Pred. NO. 0;  
Matches 1521; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

181 GTGGCGAGGATGAGATCAAGAAGAGCGTGGAGGACACGCTTGGACTTGGCTCAGTGGAC 240  
253 ACCGAAACGTCATCGATATTGTGTATCAGCCACAGGCGGTTCCTCAAGTGCAGCCAGTG 312  
241 ACCGAAACGTCATCGATATTGTGTATCAGCCACAGGCGGTTCCTCAAGTGCAGCCAGTG 300  
313 ACAAGATGACAGGATTCATCGCCGGGACAGCCGAGGCTGTGGTTCGCTGAATTCAGC 372  
301 ACAAGATGACAGGATTCATCGCCGGGACAGCCGAGGCTGTGGTTCGCTGAATTCAGC 360  
373 CCGGATGTCCTCATCTCGCCAGTGGAGTGGCGACACACACAGTGCAGTTCGGATCTT 432  
361 CCGGATGTCCTCATCTCGCCAGTGGAGTGGCGACACACACAGTGCAGTTCGGATCTT 420  
433 AACACAGAGACACCCGACTTCACCTGCACAGGTGATAGACAGTGGGTTCCTGGGTATCC 492  
421 AACACAGAGACACCCGACTTCACCTGCACAGGTGATAGACAGTGGGTTCCTGGGTATCC 480  
493 TGGGCTCCGATGGCAACGCTTGGCCAGCGGTTCGAAAGCGGCTCTATATCATCTGCG 552  
481 TGGGCTCCGATGGCAACGCTTGGCCAGCGGTTCGAAAGCGGCTCTATATCATCTGCG 540  
553 GACCCGAGACCGGTCACGACAGAGGCGGACCCCTTGGTGGGACACAGAAACATCAAC 612  
541 GACCCGAGACCGGTCACGACAGAGGCGGACCCCTTGGTGGGACACAGAAACATCAAC 600  
613 TGCTCCGCTGGGAAACCGTATCATCGGATCCGAGTGCAGGAACTTCTTCCGCCAGT 672  
601 TGCTCCGCTGGGAAACCGTATCATCGGATCCGAGTGCAGGAACTTCTTCCGCCAGT 660  
673 GGAGAGCGGACTGCGGATTTGGGACGTAAATTTGGGCGAGTGGCTTATGAACATTTGCC 732  
661 GGAGAGCGGACTGCGGATTTGGGACGTAAATTTGGGCGAGTGGCTTATGAACATTTGCC 720  
733 GGACACACAAATGCTGTGCACAGCAGTGCAGATGGGCTGGAGCGGCGCTTATTTATACATCC 792  
721 GGACACACAAATGCTGTGCACAGCAGTGCAGATGGGCTGGAGCGGCGCTTATTTATACATCC 780  
793 TCACAGATGCGACAGTGAAGATGTGGGAGCAGCTGATGGAATCTTGTGCGGACGCTC 852  
781 TCACAGATGCGACAGTGAAGATGTGGGAGCAGCTGATGGAATCTTGTGCGGACGCTC 840  
853 TCCTGGCAAGCTCACTGGGTAAACAACATTCGCTGAGCAGCCGATTTAGCTCCTCGCACT 912  
841 TCCTGGCAAGCTCACTGGGTAAACAACATTCGCTGAGCAGCCGATTTAGCTCCTCGCACT 900  
913 GGTCCATTCCATCCGGTGAAGATCGCTTCCAGAGCCAGCTCAGTTTTCAGCAGCTGAGAA 972  
901 GGTCCATTCCATCCGGTGAAGATCGCTTCCAGAGCCAGCTCAGTTTTCAGCAGCTGAGAA 960  
973 TTGAGGAAATCTGCCTTTGAAGCGCTTACAGCGGCTGTGCCCTGACGAGGTGAGTCTG 1032  
961 TTGAGGAAATCTGCCTTTGAAGCGCTTACAGCGGCTGTGCCCTGACGAGGTGAGTCTG 1020  
1033 GTTTCCTGTCGATGACACACACCTCTATCTGCGGGAACAACAACAACAGTTCGTT 1092  
1021 GTTTCCTGTCGATGACACACACCTCTATCTGCGGGAACAACAACAACAGTTCGTT 1080  
1093 GAGCGCATGACAGGCGCACAGAACGTGGTCAACGATGTGAAATATTTCCGCGGATGTAAG 1152  
1081 GAGCGCATGACAGGCGCACAGAACGTGGTCAACGATGTGAAATATTTCCGCGGATGTAAG 1140  
1153 CTAAATTCGCTGCTCTTCAATTTGACAAGTCAAGTGGTTCGCGGAGCCAGCGATGGTCAG 1212  
1141 CTAAATTCGCTGCTCTTCAATTTGACAAGTCAAGTGGTTCGCGGAGCCAGCGATGGTCAG 1200  
1213 TACATGCCACCTTCCGGGTCATGTGAGGCTGTGTACAGGTTGCTTGGTCCCGGAC 1272  
1201 TACATGCCACCTTCCGGGTCATGTGAGGCTGTGTACAGGTTGCTTGGTCCCGGAC 1260  
1273 TCCGCTTGATTTCTTCCGGCAGCAAGACCTCAACTTAAAGATGAGTGTGAGTGTGAGACG 1332





```

Db 100874 CGGACGTTCTCTGGCCACGCTCACTGGGTAAACAACATTGCGCTGAGCACGACTAGTC 100933
QY 904 CTGCGCACTGGTCCATTCATCCCGTGAAGATCGCTCAAGAGCCACCTCAGTTTGAGC 963
Db 100934 CTGCGCACTGGTCCATTCATCCCGTGAAGATCGCTCAAGAGCCACCTCAGTTTGAGC 100993
QY 964 A-----C 965
Db 100994 AGTAAGAATTATATGCTCTCTATATATGCTATATCTATAAAGAGATTGTTTACAGC 101053
QY 966 TGAGGAATTCAGGAATCTGCTTGAAGCGCTACAGCGCTGCGCCGTGAGAGGTGGA 1025
Db 101054 TGAGGAATTCAGGAATCTGCTTGAAGCGCTACAGCGCTGCGCCGTGAGAGGTGGA 101113
QY 1026 GTCGCTGTTTCTGTCGGATGACAACCCCTCTATCTGTGCGCGGAACAACACGAGAACAA 1085
Db 101114 GTCGCTGTTTCTGTCGGATGACAACCCCTCTATCTGTGCGCGGAACAACACGAGAACAA 101173
QY 1086 GTCGCTGTTGAGCGCATGACAGGCGCACAGACGTGGTCAACGATGTGAATATTCGCCGGA 1145
Db 101174 GTCGCTGTTGAGCGCATGACAGGCGCACAGACGTGGTCAACGATGTGAATATTCGCCGGA 101233
QY 1146 TGTAAAGCAATTAAGTCTGCTTCAATTTGACAAGTCAGTGTGTCGTGCGGAGCCAGCGA 1205
Db 101234 TGTAAAGCAATTAAGTCTGCTTCAATTTGACAAGTCAGTGTGTCGTGCGGAGCCAGCGA 101293
QY 1206 TGTCAGTATGATGCGCACCTTCCGGGCTCATGTGCAGGCTGTTTACAGGTTGCTGGTC 1265
Db 101294 TGTCAGTATGATGCGCACCTTCCGGGCTCATGTGCAGGCTGTTTACAGGTTGCTGGTC 101353
QY 1266 CCGGACTCCCGCTTGATGTTTCCGGCAGCAAGACTCAACTCTAA 1313
Db 101354 CCGGACTCCCGCTTGATGTTTCCGGCAGCAAGACTCAACTCTAAAGGTTAGCACGT 101413
QY 1314 -----AGTATGGAGTGTGCAGA 1330
Db 101414 AGTTTCAGTTTCTAGCGCTAGCTAATACATATTCCTCTTGTAGTATGGAGTGCAGA 101473
QY 1331 CGAAGAACTGCGACAGAGCTGCTGGACATCGCGATCGGATGTGCGAGTGTGAGTGTGG 1390
Db 101474 CGAAGAACTGCGACAGAGCTGCTGGACATCGGATGTGCGAGTGTGCGAGTGTGG 101533
QY 1391 CGCCGATGGCTCTAGAGTGTGCTGCTGGCAAGGACAAAGTTATAAGCT-----1442
Db 101534 CGCCGATGGCTCTAGAGTGTGCTGCTGGCAAGGACAAAGTTATAAGCTCTAAGTTT 101593
QY 1443 -----ATGGCGTTA 1451
Db 101594 GAGTAACGAGGTTTCTATACAAATATACAGACTCTAATGTATCTTTTCAGATGGCGTTA 101653
QY 1452 TTAACAATCATTAAGTGTACACGGTAAGAAATCTTAGGAATAAGTAAACGTCCT 1511
Db 101654 TTAACAATCATTAAGTGTACACGGTAAGAAATCTTAGGAATAAGTAAACGTCCT 101713
QY 1512 GAGTAAA 1518
Db 101714 GAGTAAA 101720

```

## RESULT 6

AC008002

LOCUS

AC008002 182726 bp DNA linear INV 06-APR-2001  
 Drosophila melanogaster, chromosome 2L, region 21D-21E, BAC clone

DEFINITION

BACR48E08, complete sequence.

ACCESSION

AC008002.3 GI:13559545

VERSION

HTG.

KEYWORDS

ORGANISM

Drosophila melanogaster.

SOURCE

Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

REFERENCE

1 (bases 1 to 182726)

## AUTHORS

Celniker, S.E., Adams, M.D., Krommiller, B., Tyler, D., Wan, K.H., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Brandon, R.C., Rogers, Y., An, H., Baldwin, D., Banzon, J., Beeson, K.Y., Busam, D.A., Carlson, J.W., Center, A., Champe, M., Davenport, L.B., Dietz, S.M., Dodson, K., Dorsett, V., Doup, L.E., Doyle, C., Dresnek, D., Farfan, D., Ferreira, S., Frise, E., Galle, R.F., Garg, N.S., George, R.A., Gonzalez, M., Houch, J., Hoskins, R.A., Hostin, D., Howland, R.J., Ibegwam, C., Jalali, M., Kruse, D., Li, P., Mattei, B., Moshrefi, A., McIntosh, T.C., Moy, M., Murphy, B., Nelson, C., Nelson, K.A., Nunoo, J., Pacleeb, J., Paragov, V., Park, S., Patel, S., Pfeiffer, B., Phouanavong, S., Pittman, G.S., Puri, V., Richards, S., Scheeler, F., Shananian, M., Strong, R., Svirska, R., Tector, C., Williams, S.M., Zaveri, J.S., Smith, H.O., Rubin, G.M. and Venter, J.C.

## TITLE

Sequencing of Drosophila chromosome 2L, region 21D-21E

Unpublished

2 (bases 1 to 182726)

Celniker, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazek, R.G.,

Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,

Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,

Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L.,

Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P.,

Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleeb, J.M., Park, S.,

Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snir, E.,

Svirska, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and

Rubin, G.M.

Direct Submission

Submitted (09-JUL-1999) Drosophila Genome Center, Lawrence Berkeley

Laboratory, MS 64-121, Berkeley, CA 94720, USA

On Apr 6, 2001 this sequence version replaced gi:6532012.

Sequence submitted by:

Berkeley Drosophila Genome Project

Berkeley National Laboratory, MS 64-121

Berkeley, CA 94720

This sequence was assembled using end sequences from a whole genome

shotgun and from subclones of this BAC and its neighboring clones.

For further information about this sequence, including its location

and relationship to other sequences, please visit our sequence

archive Web site (<http://www.fruitfly.org/sequence/>) or send emailto [bdgp@fruitfly.berkeley.edu](mailto:bdgp@fruitfly.berkeley.edu).

Location/Qualifiers

1. 182726

/organism="Drosophila melanogaster"

/strain="y; cn bw sp"

/db\_xref="taxon:7227"

/chromosome="2L"

/map="21D-21E"

/clone="BACR48E08 (D843)"

/clone\_lib="RPCI-98 (Roswell Park Cancer Institute

Drosophila melanogaster BAC library, partial EcoRI in

pBACE3.6)"

BASE COUNT 49900 a 42515 c 41297 g 49014 t

ORIGIN

Query Match 82.7%; Score 1285.8; DB 3; Length 182726;

Best Local Similarity 89.2%; Pred. No. 6.1e-311;

Matches 1505; Conservative 0; Mismatches 7; Indels 175; Gaps 4;

QY 7 CAAAAATGCGAGGACGACGAGGAGCGCGCCGCAATCGACCTGCGGAGGAATCACT 66

Db 89258 CAGAAATGCGAGGACGACGAGGAGCGCGCCGCAATCGACCTGCGGAGGAATCACT 89317

QY 67 CTCGTTT---ACACGGGCGAGGAGCGCGCCGCAATCGACCTGCGGAGGAATCACT 123

Db 89318 CTCGTTTTCGACACGCGGCGAGGAGCGCGCCGCAATCGACCTGCGGAGGAATCACT 89377

QY 124 ACCCAGCAATTGGGACTGATTTGCAACGCGCTGCTGAAAACGAGGAGCACTCCATAT 183

Db 89378 ACCCAGCAATTGGGACTGATTTGCAACGCGCTGCTGAAAACGAGGAGCACTCCATAT 89437

QY 184 TTGTTTTTCGTCGGCGAGGATGAGATCAAGAAGAGCTTGAGGAGCACCTTGGACTTGGCG 243

Db 89438 TTGTTTTTCGTCGGCGAGGATGAGATCAAGAAGAGCTTGAGGAGCACCTTGGACTTGGCG 89497

QY 244 TCAGTGGACACCGAAACGCTGATCGATATTTGTGTATCAGCCACAGCGGTTTCAAGTG 303  
Db 89498 TCAGTGGACACCGAAACGCTGATCGATATTTGTGTATCAGCCACAGCGGTTTCAAGTG 89557  
QY 304 CGCCCACTGCAAGATGCAAGATTTCCATGCCGGGACACCGAGGCTGTGGTTTCGCTG 363  
Db 89558 CGCCCACTGCAAGATGCAAGATTTCCATGCCGGGACACCGAGGCTGTGGTTTCGCTG 89617  
QY 364 AATTTACGCCCGGATGGTCTCATCTGCCAGTGAAGTGGCAGACACACAGTGGGATTG 423  
Db 89618 AATTTACGCCCGGATGGTCTCATCTGCCAGTGAAGTGGCAGACACACAGTGGGATTG 89677  
QY 424 TGGGATCTTAACACAGAGACACCGCACTTCCACTGCGACAGGTCAATAGCAGTGGGTTCTG 483  
Db 89678 TGGGATCTTAACACAGAGACACCGCACTTCCACTGCGACAGGTCAATAGCAGTGGGTTCTG 89737  
QY 484 TCGCTATCTCGGCTCGGATGCCAACGTTGGCCAGCGTTGGCAACGGGCTCTATA 543  
Db 89738 TCGCTATCTCGGCTCGGATGCCAACGTTGGCCAGCGTTGGCAACGGGCTCTATA 89797  
QY 544 ATCATCTGGGACCGGAGACGGGTGACGAGAAGGGGACCCCTTGAGTGGGCAAGAATA 603  
Db 89798 ATCATCTGGGACCGGAGACGGGTGACGAGAAGGGGACCCCTTGAGTGGGCAAGAATA 89857  
QY 604 CACATCAACTGCTCGCTGGGAACCGTATCATCGGATCGGAGTGCAGGAACCTTGCT 663  
Db 89858 CACATCAACTGCTCGCTGGGAACCGTATCATCGGATCGGAGTGCAGGAACCTTGCT 89917  
QY 664 TCCGCCAGTGAGACGGGGACTGCCGGATTTGGGACGTAAATTTGGCCAGTGCCTTATG 723  
Db 89918 TCCGCCAGTGAGACGGGGACTGCCGGATTTGGGACGTAAATTTGGCCAGTGCCTTATG 89977  
QY 724 AACATTGCGGACACAAATGCTGTGACAGCAGTGGGTTGGGAGCGGCGCTTATT 783  
Db 89978 AACATTGCGGACACAAATGCTGTGACAGCAGTGGGTTGGGAGCGGCGCTTATT 90037  
QY 784 TATACATCTCCAAAGATGCCAGTGAAGATGTGGGACGACGTGATGAATCTTGTGC 843  
Db 90038 TATACATCTCCAAAGATGCCAGTGAAGATGTGGGACGACGTGATGAATCTTGTGC 90097  
QY 844 CGGAGGTTCTCTGCCAAGCTCACTGGGTAAACACATTTGCGTGGACACCGATTACGTC 903  
Db 90098 CGGAGGTTCTCTGCCAAGCTCACTGGGTAAACACATTTGCGTGGACACCGATTACGTC 90157  
QY 904 CTGCGCACTGCTCATTCATTCGCTGAGGATGCTCCAGAGCCACCTTCAGTTTGAGC 963  
Db 90158 CTGCGCACTGCTCATTCATTCGCTGAGGATGCTCCAGAGCCACCTTCAGTTTGAGC 90217  
QY 964 A-----C 965  
Db 90218 AGTAAGAATTCATATGCTCTATATATGCTGTATATCTTAAAGAGATTGTTTACAGC 90277  
QY 966 TGAGGAATTCAGGAATCTCGCTTGAAGCGCTACAGGCGGTGTCCTGACGAGTGGA 1025  
Db 90278 TGAGGAATTCAGGAATCTCGCTTGAAGCGCTACAGGCGGTGTCCTGACGAGTGGA 90337  
QY 1026 GTGCGTGTTCCTGTCGGATGACACACCCCTCTATCTGTGGCGGAACACAGAACAA 1085  
Db 90338 GTGCGTGTTCCTGTCGGATGACACACCCCTCTATCTGTGGCGGAACACAGAACAA 90397  
QY 1086 GTGCGTTGAGCGCATGACAGGCGACAGAACGTTGCAACGATGTGAATATTTCGCCGA 1145  
Db 90398 GTGCGTTGAGCGCATGACAGGCGACAGAACGTTGCAACGATGTGAATATTTCGCCGA 90457  
QY 1146 TGTAAAGCTAATTTGGTCTGCTTCAATTTGACAAGTCAAGTGTGGCGGACCGACGA 1205  
Db 90458 TGTAAAGCTAATTTGGTCTGCTTCAATTTGACAAGTCAAGTGTGGCGGACCGACGA 90517  
QY 1206 TGTGTCAGTACATGGCCACCTTCGCGGCTCATGTGAGGCTGTTTACACGGTTCGCTGGTC 1265  
Db 90518 TGTGTCAGTACATGGCCACCTTCGCGGCTCATGTGAGGCTGTTTACACGGTTCGCTGGTC 90577  
QY 1266 CGCGGACTCCCGCTTGATTTGTTCCGGGACGAAAGACTCAACTCTAAA----- 1313

Db 90578 CGCGGACTCCCGCTTGATTGTTTCCGGCAGCAAGACTCAACTCTAAAAGCTTAGCACGT 90637  
QY 1314 -----AGTATGAGTGTGCAGA 1330  
Db 90638 AGTTTCAGTTTTCAGCGCTAGCTAATACATATCTCTTGTCTAGTATGAGTGTGCAGA 90697  
QY 1331 CGAAGAAACTGGCACAGGAGTGCCTTGGACATCGCGATGAGGTGTTTCGGAGTGGATGGG 1390  
Db 90698 CGAAGAAACTGGCACAGGAGTGCCTTGGACATCGCGATGAGGTGTTTCGGAGTGGATGGG 90757  
QY 1391 CGCCCGATGCTCTAGAGTTCCTCTGTGGCAAGACAAAGTTTATAAAGCT----- 1442  
Db 90758 CGCCCGATGCTCTAGAGTTCCTCTGTGGCAAGACAAAGTTTATAAAGCTGTAAGTTT 90817  
QY 1443 -----ATGGGCTTA 1451  
Db 90818 GAGTAACGAGGTTTCTATACATATACAGACTCTAATTTGTATCTTTCAGATGGGCTTA 90877  
QY 1452 TTAAACAATCATTAACCTTTACACGGTAAAGAAATCTTAGGAATAAAGTAAACGTCCT 1511  
Db 90878 TTAAACAATCATTAACCTTTACACGGTAAAGAAATCTTAGGAATAAAGTAAACGTCCT 90937  
QY 1512 GAGTAAA 1518  
Db 90938 GAGTAAA 90944  
RESULT 7  
AB003589/c  
LOCUS Drosophila melanogaster genomic scaffold 142000013386046 section 2  
DEFINITION of 16, complete sequence.  
ACCESSION AB003589  
VERSION AB003589.2  
KEYWORDS GI:10727410  
SOURCE HTG.  
ORGANISM Drosophila melanogaster.  
REFERENCE 1 (bases 1 to 302473)  
AUTHORS Adams,M.D., Celinker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,  
Ananidis,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galie,R.F.,  
George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,  
Sutton,G.G., Wortman,J.R., Vandeil,M.D., Zhang,Q., Chen,L.X.,  
Brandon,R.C., Rogers,Y.H., Blazei,R.G., Champe,M., Pfeiffer,B.D.,  
Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor  
Miklos,G.L., Abril,J.F., Aghayani,A., An,H.J.,  
Andrews-Pfammkuch,C., Baldwin,D., Ballew,R.M., Basu,A.,  
Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,  
Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,  
Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C.,  
Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,  
Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de  
Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,  
Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,  
Dunn,P., Durbin,K.J., Evans,J., Evangelista,C.C., Ferraz,C.,  
Fleischmann,W., Foster,C., Gabrielian,A.E., Garg,N.S., Gu,Z.,  
Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,  
Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,  
Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,  
Kensison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,  
Krivtsov,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,  
Li,J., Li,Z., Liang,X., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,  
McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,  
Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,  
Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,  
Nuskern,D.R., Pacle,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,  
Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,  
Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,  
Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,

Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R.,  
 Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wasserman,D.A.,  
 Weinstock,G.M., Weissenbach,J., Williams,S.M., Woodage,T.,  
 Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yah,R.F.,  
 Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,  
 Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,  
 Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.  
 The genome sequence of *Drosophila melanogaster*  
 Science 287 (5461), 2185-2195 (2000)  
 20196006  
 TITLE  
 JOURNAL MEDLINE  
 PUBMED  
 10731132  
 REFERENCE  
 2 (bases 1 to 302473)  
 Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.  
 Direct Submission  
 TITLE  
 JOURNAL  
 Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,  
 Rockville, MD, USA  
 COMMENT  
 On Oct 9, 2000 this sequence version replaced gi:7296179.  
 FEATURES  
 location/Qualifiers  
 1. 302473  
 /organism="Drosophila melanogaster"  
 /db\_xref="taxon:7227"  
 /chromosome="2L"  
 <3618. >39613  
 /gene="Plc21C"  
 /note="CG4574"  
 /map="21B7-21C1"  
 /db\_xref="FLYBASE:FBan0004574"  
 /db\_xref="FLYBASE:FBgn0004611"  
 join(3618..3958,11098..11376,11383..11475,19033..19179,  
 27475..27612,28770..28903,29373..29558,29704..30371,  
 30371..30553,31293..31473,32562..32796,34065..34210,  
 35933..36146,36220..37083,37396..37843,37909..38301,  
 38363..38507,38569..39613)  
 /gene="Plc21C"  
 /product="Ct14752"  
 /db\_xref="FLYBASE:FBan0004574"  
 /db\_xref="FLYBASE:FBgn0004611"  
 join(11383..11475,19033..19179,27475..27612,28770..28903,  
 29373..29558,29704..30371..30553,31293..31473,  
 32562..32796,34065..34210,35933..36146,36220..37083,  
 37396..37843,37909..38301,38363..38507,38569..38802)  
 /note="Plc21C"  
 /gene="Plc21C" gene product"  
 /codon\_start=1  
 /protein\_id="AAF51507.1"  
 /db\_xref="GI:7296215"  
 /db\_xref="FLYBASE:FBan0004574"  
 /db\_xref="FLYBASE:FBgn0004611"  
 /translation="MMSAGGYISTASVEVPOALQDGEKFIIRWDDSGTGTPVTMRVD  
 AKGFLLYVDQNNELDIIATIRDTQYAKRPDKNLKRIQIVTLGPQDLEETVT  
 VCHGSDVNTFVNCCTRRDIAQLMTDGLIKLAYSALQINGSIMEFQRAHTKLCIQ  
 VKSGRIPVKNIIKLFQAKQEDKRVEKALDVTGLPSGKVDISIVSKFQEDFYNLYK  
 YLTQRSEVERLFDISVGNKSKCMSIAQLVEFLNKTQRPRLNELIYPANPARAKEL  
 IQOYENKNAQGLSLDGLFRLMGDDNPIMAPSKLDLDDMDQPMSHVFINSNHN  
 TYLTGHQTKSSVEIRYQCLLAGRCVDELDFWNGTPEEPIVHGTYFVEIFAKDVL  
 EATAESAFKSEYPIVLSFENHCNRCQAKIANYCREIFGDMILDXPLDSHLEPNMD  
 LPPAMLRKRIIINKKKHHHHHHHHKHPAQGVTPANNKLTITANSYDARKAAQVQV  
 LSASHEGGVYTRSTANGVATGTGTSAGTAGHAPPLQQRSSKSDSTGSSDSSSS  
 EDESLPNTPLNPSGNEPPPEAKQETAGAEISALVNTVQPIHFSSFENAEKKNRCY  
 EMSFDEKQATLLKRIEPIFNYSYKLSQVYPAGTRFDSNFMPLQFLWNACCOLVA  
 LNFQTLDLAMQLNLGIFEYRNARNGYLLKFEFMRSDRLDPFAESTVDGIAGTGSIT  
 VLSQGLTKRANTEVEVDYGLPADTVRKFKTKTVRDNGMNPXYDEEPFVKVVL  
 PELASIRIAAYEGGKLGHRLVPIGLCPYGRHNLRSYGVQIALASFLCVVKVD  
 YVPDLNSFEALANPIKYQSELEKRDQLSLVDEAALGSADEDLKSCGOKKELR  
 PVESLATSPKRPSPISAAAMSDVTVTRDGDGGEFDSISIVAPSIQHSLDQSVST  
 SIRQVSSOFVDLVLAEPLEKILDHKSVEKELEMEKKLESLEKKHKKIKIAGOK  
 SSPLGKKKPAITNKLKRLSNKSLNCLSPSESGVEIPACPLDLDDGSSSESAADA  
 GEDLAGSSSLDRTQESLRKACREYISQIRELQKTHEATYSAAEKVLKTSQGTQ  
 QKLSALDKVTEVMHQLEQARNEVKNALVHRDRELIRMKREVASSVVEGRVACR  
 VRLKQTFDRTDELQKDSVRNALAPHRKAPQILDKAEARSCVSSNGFLVFLFGRP  
 HHGCTGSGSSLSGNLTLLNLDAGAGSHSAISPAKSHNSTAAAEWKT"  
 complement(<11712. >12601)  
 /gene="CG11912"

/map="21B7-21B8"  
 /db\_xref="FLYBASE:FBan0011912"  
 /db\_xref="FLYBASE:FBgn0031248"  
 complement(11712..12601)  
 /gene="CG11912"  
 /product="CT37080"  
 /db\_xref="FLYBASE:FBan0011912"  
 /db\_xref="FLYBASE:FBgn0031248"  
 complement(11773..12588)  
 /gene="CG11912"  
 /note="CG11912 gene product"  
 /codon\_start=1  
 /protein\_id="AAF51509.1"  
 /db\_xref="GI:7296217"  
 /db\_xref="FLYBASE:FBan0011912"  
 /db\_xref="FLYBASE:FBgn0031248"  
 /translation="MKQFAVIFALALASVSAISVQPGPFPEGRINYEAAKEAEPYI  
 VSIQTTNSHFCAGSLDELVTITAAHCLTYNQGAAGHSRTDQENVQIRKFTNAQ  
 YVIHENYGGVGPNDIGLILKEEDAFDLNVAADGSPNVAISVSKPTFGTSDGYL  
 YGWRDNGSLPLNLQKLDALIVDYNECKAALFSPNNSLAETNVCTHTPGKADSCNGD  
 SGGPLVSQSSRGAELIGIVSWGYPCLSTTYPVSVTVSSFLPWIDENRKA"  
 complement(<13648. >14580)  
 /gene="CG11911"  
 /map="21B7-21B8"  
 /db\_xref="FLYBASE:FBan0011911"  
 /db\_xref="FLYBASE:FBgn0031249"  
 complement(13648..14580)  
 /gene="CG11911"  
 /product="CT37076"  
 /db\_xref="FLYBASE:FBan0011911"  
 /db\_xref="FLYBASE:FBgn0031249"  
 complement(13743..14576)  
 /gene="CG11911"  
 /note="CG11911 gene product"  
 /codon\_start=1  
 /protein\_id="AAF51508.1"  
 /db\_xref="GI:7296216"  
 /db\_xref="FLYBASE:FBan0011911"  
 /db\_xref="FLYBASE:FBgn0031249"  
 /translation="NKLITVFLVIALVAAQAQAKLSKLKLVPSFATGFVINGTEAE  
 PHAPATIVSLATNLYLKHSHICGGTLINKOMIVTAACHISEPVGMISITAGLTHAEVDE  
 LTKOORVDGRVREKTVGCVGYDIALLVHVNESFIFENWOPATLPGRQOVHEGETHL  
 YGMGPKSVIFSSAKTLQTVTQILNVECKEELPESAPIAESNICSSSLQSKSACN  
 GDSGGPLVVEFTNAPSELIGIVSWGYPCLGLANKPFIYKVSAYDWTNIQSYKYL  
 F"  
 complement(<49281. >51230)  
 /gene="CG11907"  
 /map="21C1-21C1"  
 /db\_xref="FLYBASE:FBan0011907"  
 /db\_xref="FLYBASE:FBgn0031250"  
 complement(join(49281..49459,49730..50200,50254..50554,  
 50786..51230))  
 /gene="CG11907"  
 /product="CT3136"  
 /db\_xref="FLYBASE:FBan0011907"  
 /db\_xref="FLYBASE:FBgn0031250"  
 complement(join(49281..49459,49730..50200,50254..50554,  
 50786..51049))  
 /gene="CG11907"  
 /note="CG11907 gene product"  
 /codon\_start=1  
 /protein\_id="AAF51506.1"  
 /db\_xref="GI:7296214"  
 /db\_xref="FLYBASE:FBan0011907"  
 /db\_xref="FLYBASE:FBgn0031250"  
 /translation="MPCSTDFERLVADLDDDEILTGEGSVFAPRQDHHADQCLM  
 ERQDEEVLVSNPEPSGRFLTYIVYLLGIGMTPNFHVTAEDYWKYKFNASINNT  
 DLEEEUPLQKSTCDLALTATISGTTFFLLNLAIFGHVSVTRKMLGTLMMILIFGV  
 TGFVINTDITWQEFLLITLIVLLNLSAATMSGALYVAGLVFSEFTATVYVSOQA  
 LGGLIFALFIVLAFDTPNTATIFFYVGVLLICLVIVILAKRPFYVLEGG  
 DKYKTRAVPSHNRSAGEPLERILROVMSKIVLHAIISALLYTTLTSLVPAVTVL  
 MOSEYGHVSPTAMVMYFALSNGYFTNILLTMAPKRVKQHEKELASSITMAAALSCGNV  
 GSLLSLVFVQML"



gene	<68682..>77995 /gene="al" /note="CG3935" /map="21C1-21C2" /db_xref="FLYBASE:FBan0003935" /db_xref="FLYBASE:FBgn0000061" join(68682..69037,75067..75450,76257..76302,76864..77132,77259..77995) /gene="al" /product="CT12963" /note="Nucleotide sequence of the Celera sequence differs		
mRNA	Query Match Best Local Similarity 82.7%; Score 1285.8; DB 3; Length 302473; Matches 1505; Conservative 0; Mismatches 7; Indels 175; Gaps 4;  7 CAAAAATGCGAGACGACGACGAGCAAGAGCGCCACGCCACATACGATACAGGCGCGC 66 258606 CAGAAATGCGAGACGACGACGAGCAAGAGCGCCACGCCACATACGATACAGGCGCGC 258547  67 CTCGTTT--ACACGGGCGAGGAAGCGGCCGCCCAATCGACCTGCCGCGAGGAATCACT 123 258546 CTCGTTTCGACACGCGGCGAGGAAGCGGCCGCCCAATCGACCTGCCGCGAGGAATCACT 258487  124 ACCCAGCAATGGGACTGATTGGCAACGCGCTGCTGAAAAACGAGGAGCACTCCATAT 183 258486 ACCCAGCAATGGGCTGATTGGCAACGCGCTGCTGAAAAACGAGGAGCACTCCATAT 258427  184 TTGTTTTTCGCGGCGAGGATGAGATCAAGAAGAGCTGGAGGACAGTTGGACTTGGCG 243 258426 TTGTTTTTCGCGGCGAGGATGAGATCAAGAAGAGCTGGAGGACAGTTGGACTTGGCG 258367  244 TCAGTGACACCGAANAACGTGATCGATATTGTGTATCAGCCACAGCGGGTTTCAAGTG 303 258366 TCAGTGACACCGAANAACGTGATCGATATTGTGTATCAGCCACAGCGGGTTTCAAGTG 258307  304 CGCCAGTCACAAGATGCACGAGTTCCATCGCGGGACACGCCAGGCTGTGGTTTCGCTG 363 258306 CGCCAGTCACAAGATGCACGAGTTCCATCGCGGGACACGCCAGGCTGTGGTTTCGCTG 258247  364 AATTTCAGCCCGGATGGTCTCATCTCGCCAGTGGAGTGGCGACACACAGTGCATATG 423 258246 AATTTCAGCCCGGATGGTCTCATCTCGCCAGTGGAGTGGCGACACACAGTGCATATG 258187  424 TGGATCTTAACACAGACACCGCACTTCACCTGCACAGGTCTATAAGCAGTGGTCTG 483 258186 TGGATCTTAACACAGACACCGCACTTCACCTGCACAGGTCTATAAGCAGTGGTCTG 258127  484 TGCATATCTTGGCTCCGGATGGCAACGGTTGGCCAGCGTTGCAAGCGGCTCTATA 543 258126 TGCATATCTTGGCTCCGGATGGCAACGGTTGGCCAGCGTTGCAAGCGGCTCTATA 258067  544 ATCATCTGGACCCCGAGACGGTTCACAGAAAGGGCGACCCCTTGAGTGGGCAACAGAA 603 258066 ATCATCTGGACCCCGAGACGGTTCACAGAAAGGGCGACCCCTTGAGTGGGCAACAGAA 258007  604 CACATCAACTGCTCGCTCGCTGGGAAACCGTATCATCGCGATCCGGAGTGCAGGAACTTGC 663 258006 CACATCAACTGCTCGCTCGCTGGGAAACCGTATCATCGCGATCCGGAGTGCAGGAACTTGC 257947  664 TCCGCCAGTGGAGACGGGACTGCCGATTTGGAGCTTAATAATTTGGCCAGTGCCTTATG 723 257946 TCCGCCAGTGGAGACGGGACTGCCGATTTGGAGCTTAATAATTTGGCCAGTGCCTTATG 257887  724 AACATTGCCGGACACACAAATGCTGTGACAGCAGTGGAGTGGGTGGAGCGGCCCTTAT 783 257886 AACATTGCCGGACACACAAATGCTGTGACAGCAGTGGAGTGGGTGGAGCGGCCCTTAT 257827  784 TATACATCTTCCAAAGATCGCACAGTGAAGATGTGGCGAGCAGTGTGAATCTTTGTC 843 257826 TATACATCTTCCAAAGATCGCACAGTGAAGATGTGGCGAGCAGTGTGAATCTTTGTC 257767  844 CGGACGTTCTCTGGCCAAAGCTCACTGGGTAACAACATTTGGCTGAGCAGCCGATFAGTC 903		
		Db	257766 CGGACGTTCTCTGGCCACGCTCACTGGGTAACAACATTTGGCCAGCAGTGTGAATCTTTGTC 257707
		QY	904 CTGCGCACTGGTCCATTCATCCGTTGAAGGATCGCTCCAAGAGCACCTCACTGTTTGAAGC 963
		Db	257706 CTGCGCACTGGTCCATTCATCCGTTGAAGGATCGCTCCAAGAGCACCTCACTGTTTGAAGC 257647
		QY	964 A-----C 965
		Db	257646 AGAAGAATTCATATGCTCTATATATGCTGTATATCTCTAAAGAGATGTTTACAGC 257587
		QY	966 TGAGGAATTCAGGAATTCGCTTTGAAGCGCTACCAAGCGCTGTGCCCTGACGAGTGGA 1025
		Db	257586 TGAGGAATTCAGGAATTCGCTTTGAAGCGCTACCAAGCGCTGTGCCCTGACGAGTGGA 257527
		QY	1026 GTGCGTGGTTTCTGCTCGGATGACAAACCCCTCTATCTGTGGCGGAACACAGAACAA 1085
		Db	257526 GTGCGTGGTTTCTGCTCGGATGACAAACCCCTCTATCTGTGGCGGAACACAGAACAA 257467
		QY	1086 GTGCGTTGAGCGCATGACAGGGACAGAAAGCTGTGTCAACGATGTGAATATTTCCGCGGA 1145
		Db	257466 GTGCGTTGAGCGCATGACAGGGACAGAAAGCTGTGTCAACGATGTGAATATTTCCGCGGA 257407
		QY	1146 TGTAAAGCTAATTGCGCTGCTTCATTGTGACAGTCAAGTGCCTGTGTGGCGAGCCAGCGA 1205
		Db	257406 TGTAAAGCTAATTGCGCTGCTTCATTGTGACAGTCAAGTGCCTGTGTGGCGAGCCAGCGA 257347
		QY	1206 TGTTCAGTACATGGCCACCTTCCGGGGTCAATGTGAGGCTGTGTACAGGTGCTTGGTC 1265
		Db	257346 TGTTCAGTACATGGCCACCTTCCGGGGTCAATGTGAGGCTGTGTACAGGTGCTTGGTC 257287
		QY	1266 CGCGGACTCCGCTTGATGTTTCCGGCAGCAAGACATCAACTCTAAA-----1313
		Db	257286 CGCGGACTCCGCTTGATGTTTCCGGCAGCAAGACATCAACTCTAAAAGGTTAGCACCT 257227
		QY	1314 -----AGTATGGAGTGTGCAGA 1330
		Db	257226 AGTTTCAGTTTTCTAGCGCTAGCTAATACATATCCCTCTTGTGCTAGTATGGAGTGTGCAGA 257167
		QY	1331 CGAAGAACTGCGACAGGAGCTGCTGGACATGCGGATGCGGATGCGGAGTGCAGTGG 1390
		Db	257166 CGAAGAACTGCGACAGGAGCTGCTGGACATGCGGATGCGGATGCGGAGTGCAGTGG 257107
		QY	1391 CGCCGATGGCTTAGAGTTCGCTGTGGTGGCAAGACAAAGTTATATAAGCT-----1442
		Db	257106 CGCCGATGGCTTAGAGTTCGCTGTGGTGGCAAGACAAAGTTATATAAGCTTAAAGCTT 257047
		QY	1443 -----ATGGGCTTA 1451
		Db	257046 GAGTAACGAGGTTTTCTATATAACATATACAGACTCTAATTTGTATCTTTTCAGATGGGCTTA 256987
		QY	1452 TTAAACAATCATTAATTTACGCTGACGCTGACGCTGACGCTGACGCTGACGCTGACGCT 1511
		Db	256986 TTAAACAATCATTAATTTACGCTGACGCTGACGCTGACGCTGACGCTGACGCTGACGCT 256927
		QY	1512 GAGTAAA 1518
		Db	256926 GAGTAAA 256920
RESULT 8			
AX135811			
LOCUS	AX135811	1859 bp	DNA linear PAT 29-MAY-2001
DEFINITION	Sequence 5 from Patent WO0132614.		
ACCESSION	AX135811		
VERSION	AX135811.1	GI:14272046	
KEYWORDS			
SOURCE	human.		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
REFERENCE	1 (bases 1 to 1859)		
AUTHORS	Utku.N.		





TITLE Nakamura,Y., Nagahari,K., Masuho,Y., Ninomiya,K. and Iwayanagi,T.  
 JOURNAL NEDO human cDNA sequencing project  
 REFERENCE Unpublished  
 AUTHORS 2 (bases 1 to 1853)  
 TITLE Isogai,T. and Otsuki,T.  
 JOURNAL Direct Submission

COMMENT Submitted (16-FEB-2000) Takao Isogai, Helix Research Institute,  
 Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
 (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)  
 NEDO human cDNA sequencing project supported by Ministry of  
 International Trade and Industry of Japan: cDNA full insert  
 sequencing: Research Association for Biotechnology; cDNA library  
 construction, 5'- & 3'-end one pass sequencing and clone selection:  
 Helix Research Institute (supported by Japan Key Technology Center  
 etc.) and Department of Virology, Institute of Medical Science,  
 University of Tokyo.

FEATURES Location/Qualifiers

1..1853

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="NT2RP1001457"

/cell\_line="NT2"

/cell\_type="teratocarcinoma"

/clone\_lib="NT2RP1"

/note="cloning vector: pUC19FL3-mRNA from NT2 neuronal  
 precursor cells after 48-hours retinoic acid (RA)  
 induction."

30..1487

/note="unnamed protein product"

/codon\_start=1

/protein\_id="BA091621.1"

/db\_xref="GI:7022502"

/translation="MAAAVADENAVRDVORLLVQFQDREGGLGSPFVDPVDTDPRL  
 QLYCNALLAQSDPLPLAFFVHDAEIVSSIGKTLAESQAVETEKVDIIYQOAFIRVRA  
 VTRCSYSLLEGHSEAVISVAFPTGKYLAGSGDFTVRFWDLSTETPFTCKGHRHWL  
 SWSGPDGKLKASCKNGQILLWDPSTGKQVGRITLHSGWITGLSWEPLHNPICRY  
 VASSKDGSRVINDTAGCERILITGHTQVTCFLRWGDDGLLYSASDRITIKVWAHD  
 GVLCTRLQGHGHWNTMALSTYALRTGAFPAEASVNPDLQSLQELKERALSRYN  
 LVYGGPERLVSGDDFTFLFWSPADKKPLTRMTGHQALINOVLPSPSRIVASNSF  
 DKSIKLWDRGTGKYLASLRGHVAAYQIWAASDSELLVSGSSDSTLKVWDVKAQKLAM  
 DLPGHADREYAVDMSPDQORVASGGKDKLRWR"

BASE COUNT 378 a 549 c 565 g 361 t

Query Match 30.7%; Score 477.4; DB 9; Length 1853;  
 Best Local Similarity 59.8%; Pred. No. 1.2e-108;  
 Matches 837; Conservative 0; Mismatches 556; Indels 6; Gaps 2;

QY 54 GATACAGCGCGCCTCGTTTACACGGCGGAGGAAGCGCGCGCAATCGACCTGCCGCG 113  
 DB 80 GCTAGTGCAATTCAGGATGAGGCGGCGAGCTGCTGGGTTCCCGGTTCCAGCTGCCGT 139  
 QY 114 AGGAATCACTACCCAGCAATTTGGACATGATTTGCAACGGGCTGCTGAAACAGGAGAC 173  
 DB 140 GGACATCAACCGCGGACAGGCTGCAGCTGCTGTGCAACGGCTACTGCGCCAGGAGATCC 199  
 QY 174 CACTCATATTTGTTTTCGTGGCGGAGGATGAGATCAAGAAGAGCGCTCGGAGGACACGTT 233  
 DB 200 CTGCGCACTGGCTTTCTTGTCCAGCATGCTGAGATGCTCTCTCTCATCTGGGGAGAGCTT 259  
 QY 234 GGACTTGGGCTGAGTGCACACCAACGATGATGATTTGTATATCAGCCACAGCGGCT 293  
 DB 260 GGATCCCGAGGCTGGAGACAGAGGTTCTAGACATCATCTACCAGCCACAGGCTAT 319  
 QY 294 TTTCAGTGGCCCGGAGTGAAGATGCACAGTTCCTCATGCGGGGACACCGCAGGCTGT 353  
 DB 320 CTTCAAGTCCGGGCTGTGACTCGCTGCACAGCTCTCTTGGAGGTTCAAGTGAAGCACT 379  
 QY 354 GGTTCGCTGAATTCAGCCCGGATGCTCATCTCCCGAGTGGAGTGGCGACAC 413  
 DB 380 CATTTCTGTGGCTTCAGCCCTACGGGAAGATACCTGGCCAGTGGCTCTGGAGACAC 439  
 QY 414 AGTGGCATGTGGGATCTTAACACAGAGACACCGCACTTCACCTGCACAGGTATAAGCA 473

DB 440 CGTGGCTTCTGGGATCTCAGACAGAGACACCATTTTCATATGCAGGAGCACAGACA 499  
 QY 474 GTGGGTTCTGCGTATCTCTGGGCTCCGGATGGCAACGGTTTGGCCAGCGGTTGCAAGC 533  
 DB 500 CTGGGTCTTATAGTATCTCTGGTCTCCAGATGGCAAGAGCTGCGCTCAGGCTGCAAGA 559  
 QY 534 GGGCTCTATAATATCATCTGGACCCCGGAGACGGTCCAGCAAGGCGGCGACCTTGAGTGG 593  
 DB 560 TGGCCAGATTTCTCTCTGGGACCCCAAGCACAGGAGACGGTGGGAGGACCTCTCGCTGG 619  
 QY 594 GCACAAGAAACATCATCACTGCTCCCTCGGGAACCGGTATCATCGGATCCGGAGTGCAG 653  
 DB 620 CCACAGCAAGTGGATCAGAGGCTGAGCTGGGAGCCCTCCATCGCAACCTGAGTGGCG 679  
 QY 654 GAAACTTCTTCCGCCAGTGGAGCGGGAGCTGCCGATTTGGGACGATAAATTTGGGCA 713  
 DB 680 CTATGTGCCAGCAGCTCCAAAGATGGCAGTGTGGGATCTGGGACACAACTTCAGGCGG 739  
 QY 714 GTGCCCTTATGAACATTTCCCGGACACACAAATGCTGTGACAGCAGTGTGAGTGGGTTGAG 773  
 DB 740 CTGTGAGGCGCATCTCTACCGGGCACACCCAGTCGGTCACTGTCTCCGTGGGAGGGA 799  
 QY 774 GGGCTTATTTATACATCTCCAAAGATCGCACAGTGAAGATGTGGCGAGCAGCTGATGG 833  
 DB 800 CGGGCTTCTCTACTCTCTCCAGGACCGCACCATCAAAAGTCTGGAGAGCTCATGACGG 859  
 QY 834 AATCTGTGGCGGACGTTCTTGGCCAGCTCAGCTGGGTAAACAATTCGCTGAGCAGAC 893  
 DB 860 TGTGCTGTGGCGGACTCTGAAGGCCACGGCCACTGGGTGAACACCATGCCCTCAGCAC 919  
 QY 894 CGATTACGCTCTCGGCACCTGGTTCATTCATCCGTTGAAGGATCGCTCCCAAGAGCC--A 950  
 DB 920 TGACTATGCTCTCGGCACCTGGGCGCTTTGAACCTGCTGAGGCTCAGATTAATCCCAAGA 979  
 QY 951 CCTCAGTTTGGAGCAGTGAAGATTCGAGGAATCTGCCTTGAAGCGCTACACGCGCTGTG 1010  
 DB 980 CCTCAAGGATCTTTCAGAGATTGAAGAGAGGGCTCTGAGCCGATACAACTCGTGGC 1039  
 QY 1011 CCCTGACAGGTGAGTTCGCTGTTTCTGTTTCGGATGACCAACACCTCTATCTGTGG-- 1068  
 DB 1040 GGGCCAGGGTCCAGAGAGGCTGTGTCTGGCTCGGAGGACTTACCTTATCTCTGTGTC 1099  
 QY 1069 -CGGAACAACACAGCAAGTGGGTTGAGCGCATGACAGGCGACAGCAAGCTGGTCAACGA 1127  
 DB 1100 CCCAGCAGGACAAAAAGCCCTCTCACTCGGATGACAGGACACCAAGCTCTCATCAACA 1159  
 QY 1128 TGTGAATATTCGCCGGATGTAAGCTAATTCGCTCTGCTTCAATTTGACAACTGAGTGGC 1187  
 DB 1160 GGTGCTCTCTCTCTGACTCCCGCATCGTGGCTAGTGTGCTCTCTTTGACAAGTCCATCA 1219  
 QY 1188 TCTGTGGCAGCGCAGGATGTCAGTACATGGCCACCTTCCGGGGTCAATGTGACAGGCTGT 1247  
 DB 1220 GCTGTGGGATGGCAGCGGCAAGTACCTGGCTTCCCTACGGGCGACGTTGGTGGCTGTG 1279  
 QY 1248 TTACACGTTGCTGCTGCGGACTCCCGCTTGAATTTGTTCCGGCAGCAAGACTCAAC 1307  
 DB 1280 GTACCAAGATTTCGCTGAGTGTGACATGTCGGCTCCTGGTCAGGCGGACAGTGAAGCAG 1339  
 QY 1308 TCTAAAAAGTATGAGTGTGTCAGACAGAAAGAACTGGCAGAGAGCTGCTTGGACATGCGGA 1367  
 DB 1340 ACTGAAGTGTGGGATGTGAAGGCCCAAGAGCTGGACCTGGACCTGCCCGGCGACGCGGA 1399  
 QY 1368 TGAGGTGTTCGGAGTGGGCGCCGCGGCTGTAGAGTGTGCTCTGAGTGTGCTCTGTTGGCAGGA 1427  
 DB 1400 TGAGTATATGCTGTGATGCTGGAGTCCAGATGGCCAGAGTGGCAAGTGGTGGGAGGA 1459  
 QY 1428 CAAGATTATTAAGCTATGG 1446  
 DB 1460 CAATGCTCCGATATGG 1478

RESULT 10



**COMMENT**

Contact: MGC help desk  
Email: cgabps-r@mail.nih.gov  
Tissue Procurement: DCTD/DTP  
cDNA Library Preparation: Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: National Institutes of Health Intramural  
Sequencing Center (NISC),  
Gaithersburg, Maryland;  
Web site: <http://www.nisc.nih.gov>  
Contact: [nisc.mgc@nih.gov](mailto:nisc.mgc@nih.gov)  
Shevchenko, Y., Wetherby, K.D., Beckettrom-Sternberg, S.M.,  
Benjamin, B., Blackley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,  
Dierich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R.,  
Lim, M., Maduro, Q.L., Mastello, C., Mastrian, S.D., McClostey, J.C.,  
McDowell, J., Pearson, R., Snyder, B., Stantirpop, S., Thomas, P.J.,  
Tsong, L.E., Touchman, J.W., Tsurgeon, C., Vogt, J.L., Walker, M.A.,  
Zhang, L.-H. and Green, E.D.

	BASE COUNT	ORIGIN
1	0	0
2	0	0
3	0	0
4	0	0
5	0	0
6	0	0
7	0	0
8	0	0
9	0	0
10	0	0
11	0	0
12	0	0
13	0	0
14	0	0
15	0	0
16	0	0
17	0	0
18	0	0
19	0	0
20	0	0
21	0	0
22	0	0
23	0	0
24	0	0
25	0	0
26	0	0
27	0	0
28	0	0
29	0	0
30	0	0
31	0	0
32	0	0
33	0	0
34	0	0
35	0	0
36	0	0
37	0	0
38	0	0
39	0	0
40	0	0
41	0	0
42	0	0
43	0	0
44	0	0
45	0	0
46	0	0
47	0	0
48	0	0
49	0	0
50	0	0
51	0	0
52	0	0
53	0	0
54	0	0
55	0	0
56	0	0
57	0	0
58	0	0
59	0	0
60	0	0
61	0	0
62	0	0
63	0	0
64	0	0
65	0	0
66	0	0
67	0	0
68	0	0
69	0	0
70	0	0
71	0	0
72	0	0
73	0	0
74	0	0
75	0	0
76	0	0
77	0	0
78	0	0
79	0	0
80	0	0
81	0	0
82	0	0
83	0	0
84	0	0
85	0	0
86	0	0
87	0	0
88	0	0
89	0	0
90	0	0
91	0	0
92	0	0
93	0	0
94	0	0
95	0	0
96	0	0
97	0	0
98	0	0
99	0	0
100	0	0

Query Match	30.7%	Score 477.4;	DB 9;	Length 1870;
Best Local Similarity	59.8%;	Pred. No. 1.2e-108;		
Matches 837;	Conservative	0;	Mismatches 556;	Indels 6;
				Gaps 2

QY	54	GATACAGCGCGCGCTCGTTTTACAGGGCGAGGAAGCGGCCGCCCAATGACCTCGCCGGC	113
Db	58	GCTAGTCAGTTCCAGGATGAGCGGGGACGTGCTGGGTTCGCCGTTCGACGTGCCCGT	117
QY	114	AGGAATCACTACCCACCAATTGGGACTGATTTGCAACGCGCTGCTGAAAACGAGGAAGC	173
Db	118	GGACATCACCCCGGCAGACGCTGCAGCTCGTGTGCAACGCGCTACTGCCCCAGGAGATCC	177
QY	174	CACCTCCATATTGTGTTTTTCGTGGCGAGGATGAGATCAAGAAGAGCCTCGGAGACAGTT	233
Db	178	CTTGCCACTGGCTTCTTTGTGCCAGATGCTGAGATCGCTCCTCAGTGGGAGAGAGTT	237
QY	234	GGACTTTGGCGTCGATGGGACACCGAAAAGTGATCGGATTTGTGTATACGCCACAGGCGGT	293
Db	238	GGAGTCCCGACGAGTCGGACACAGAGAAGGTCCCTAGACATCATCTTACACGCCACAGGTAT	297
QY	294	TTTCAAGTGGCGCCGAGTGACAAAGATGCAGAGTTTCCATGCGCGGAGACAGCCGAGGCTGT	353

QY 1428 CAAAGTTAAAGCTATGG 1446  
 DB 1438 CAAATGCTCCGGATATGG 1456  
 RESULT 12  
 LOCUS BC012075 2593 bp mRNA linear PRI 06-AUG-2001  
 DEFINITION Homo sapiens, hypothetical protein similar to beta-transducin family, clone MGC:19970 IMAGE:4652713, mRNA, complete cds.  
 ACCESSION BC012075  
 VERSION BC012075.1 GI:15082334  
 KEYWORDS MGC.  
 SOURCE Homo sapiens.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 2593)  
 Strausberg, R.  
 Direct Submission  
 Submitted (02-AUG-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA  
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 Contact: MGC help desk  
 Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: Rubin Laboratory  
 cDNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)  
 BC Cancer Agency, Vancouver, BC, Canada  
 info@bcsc.bc.ca  
 Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield, Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin, Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven Ness, Fawn Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline Schein, Duane Smalius, Michael Smith, Lorraine Spence, Jeff Stott, Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Marra.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Series: IRAL Plate: 29 Row: O Column: 21  
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 7022501.  
 FEATURES  
 location/Qualifiers  
 1..2593  
 /organism="Homo sapiens"  
 /db\_xref="LocusID:54475"  
 /db\_xref="taxon:9606"  
 /clone="MGC:19970 IMAGE:4652713"  
 /tissue\_type="Placenta, choriocarcinoma"  
 /clone\_lib="NIH\_MGC\_21"  
 /lab\_host="DH10B-R"  
 /note="vector: pOTB7"  
 14..1471  
 /codon\_start=1  
 /product="hypothetical protein similar to beta-transducin family"  
 /protein\_id="AAH12075.1"  
 /db\_xref="GI:15082335"  
 /translation="MAAAVADEAVARDVORLLVQFQDEGGQLGSPFDPVDPDIIPDRLLVNCNALLAGDPLPLAFFVDAEIVSLGKLTLESQAVTEKVLDIIYQPAIFRVRVAVTRCTSLSEGHSEAVISVAFPTGYLAGSGDPTVRFWDLSTETPHFTCKGRHWVLSTWSGPKKLASCKNGOILLNDPSTCKQVGTLAGHSKWTGTSWEPLHNDPECHVAVSSKDGSVRIINDTTFAGRCERILGHQSVTCGLRWGGDCLLYSASORTIKVRAHDGVLCRTLOGHHWNWMTALSTDYALTKGAFPAEASVNFQDQGSQELKERRLRNLYRQSGDRLVSGDDFTFLFWSPAEDKKPLRTMTGHOALINQVLFPSPDSRIVASAFDKSKLMDRGRTGYLASLRGHVAAYQIAWSADSRLLVSGSSDSTLKVMYDAQKLIAMDLPGHAEVYAVDWSPDGQVAVSGGKDKLRINRR"  
 CDS

BASE COUNT	582 a	734 c	749 g	528 t	
ORIGIN					
Query Match	30.7%	Score 477.4;	DB 9;	Length 2593;	
Best Local Similarity	59.8%;	Pred. No. 1.3e-108;			
Matches	837;	Conservative 0;	Mismatches 556;	Indels 6;	Gaps 2;
QY	54	GATACAGCGCGCTCGCTTTTACACAGGGGAGGAGGAGCGGCGCCGCAATCGACCTGCCCGC	113		
DB	64	GTAGTGGAGTTCAGAGATGAGGGCGGCGAGCTGCTGGTTCCTCGAGCTGCCCGT	123		
QY	114	AGGAATCACTACCCAGCAATTTGGAGTGAATTTGCAACGCGCTGCTGAAAAACGAGGAGC	173		
DB	124	GGACATCACCCCGGACAGGCTGAGCTGCTGTGCAACGCGCTACTGGCCACGAGGATGCC	183		
QY	174	CACCTCATATTTGTTTTTTCGTGGGCGAGGATGAGATCAAGAAGAGCTCTGAGGACACGTT	233		
DB	184	CTGCGCACTGGCTTTCTTCTCCACGATGCTGAGATGCTCTCTCACTGGGAGACGCTT	243		
QY	234	GGACTTGGCGTCACTGGACACCGAAACGTCATGATGATATTGTGTATCACCACACGGCGT	293		
DB	244	GGAGTCCCGAGCTGGAGACAGAGAAGCTCTAGACATCATCTACCACGACAGGCTAT	303		
QY	294	TTTCAAGTGGCGCCAGTGAAGATGACAGATTCCATGCGGGGACACGCCGAGGCTGT	353		
DB	304	CTTCAGAGTCCGGGCTGTGACTGCTGTCACACAGCTCTTGGAGGGTCAACAGTGAAGGCA	363		
QY	354	GGTTTCGCTGAATTTTCAGCCCGGATGGTGTCTCATCTCGCCAGTGGAGTGGGACACAC	413		
DB	364	CATTTCTGCGCTTACGCCCTACGGGAAAGTACCTGGCCAGTGGCTCTGGAGACACAC	423		
QY	414	AGTGGGATTTGGGATCTTAACACAGACACCGCACTTCACCTGCACAGGTCATAAGCA	473		
DB	424	CGTGGCTCTGGGATCTCAGACAGACACACCATTTTACATGCAAGGACACAGACA	483		
QY	474	GTGGGTCTGTGCGTATCTCTGGGCTCCGGATGGCAACGGTTGGCCAGCGTTGCAAGC	533		
DB	484	CTGGGTCTCTAGTATATCTCTGCTCCAGATGGCAAGAGCTGGCTCTAGGCTGCAAGAA	543		
QY	534	GGGCTCTATAATCATCTGGGACCGGAGACGGGTGACGAGAGGGCGGACCCCTTGAGTG	593		
DB	544	TGCGCAGATTCTCTCTGGGACCCAAAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	603		
QY	594	GCACAGAAACACATCAACTCTCGCTCGGGAACCGTATCATCGATGCTCCGAGTCCGAG	653		
DB	604	CCACAGCAAGTGGATCAGACGCTGAGCTGGGAGCCCTCCATCGCAACCCCTGAGTGGC	663		
QY	654	GAACTTGTCTCCGCGAGTGGAGACCGGGACTGCGCGATTTGGGACGTAAATWTGGCCA	713		
DB	664	CTATGTGGCCAGCAGCTCCAAAGATGGGAGTGTGGGATCTGGGACACACACTGCAGGCG	723		
QY	714	GTGCGCTTATGAACATTTGCCGGACACACAAATGCTGTGACAGCAGTGAAGTGGGGTGGAGC	773		
DB	724	CTGTGAGCGCATCTCTACCGGGCACACCCAGTCCGCTACCTGTCTCGGTTGGGAGGGA	783		
QY	774	GGCGCTTATTTATATCTCCAAAGATCGACAGCTGAAGATGTGGGAGCAGCTGATGG	833		
DB	784	CGGGTCTCTCTCTCTCTCCCGAGGCCACCATCAAGTCTGGAGAGCTCATGACGG	843		
QY	834	AATCTTGGCGGAGTCTCTCTGGCCAAAGCTCATCGGTGAAGGATGCTCTCAAGAGCC	893		
DB	844	TGTGTGTGGCGGAGCTCTGCAAGGCCACCGGCCACTGGGTGAACACCATGGCCCTCAGC	903		
QY	894	CGATTACGCTCTGGCGCACTGGTCCATTCCTCGGTGAAGGATGCTCTCAAGAGCC	950		
DB	904	TGACTATGCCCTGGCGCACTGGGCGCTTTGAACCTCTGAGGCGCTAGTTAATCTCCCAAGA	963		
QY	951	CTCAGTTTGGAGCACTGAGCAATTTGAGCAATCTGCTCTGAAGCGTACACGCGCTGTG	1010		
DB	964	CTTCAAGATCCTTGCAGAGGTGAAGGAGGAGGCTCTGAGCGGATACACACTCTGTGG	1023		
QY	1011	CCCTGACGAGTGGAGTGGCTGCTTCTGTGTTGGATGACACACACCTCTATCTGTGG	1068		

Db 1024 GGGCAGGGTCCAGAGAGGCTGTCTGTGCTCCGACGACTTACCTTATTCTTGTGGTC 1083  
QY 1069 -CGGAACAACAGACAAGTGGTTGAGCGATGACAGGCGACACAGAGTGTGTCACGA 1127  
Db 1084 CCACGAGAGGACAAAGAGCCTCTCAGTCGGATGACAGGACCAAGCTCTCATCAACCA 1143  
QY 1128 TGTGAATATATTCGCGGATGTAAAGCTAAATTTGCTCTGTCTTTCATTTGACAAGTCAGTGG 1187  
Db 1144 GTGTCTCTCTCTCTGACTCCCGCATCGTGGCTAGTGGCTCTCTTTGACAAGTCCATCAA 1203  
QY 1188 TCTGTGGCGAGCGAGCGATGGTCAGTACATGGCCACCTTCCGGGTGATGTGACGGCTGT 1247  
Db 1204 GCTGTGGGATGGCAGGACGGCAAGTACCTGGCTTCCCTACGGCCAGTGGCTGGCGT 1263  
QY 1248 TTACAGCTGTGGTGGTCCGCGACTCCCGCTGATGTTTCCGCGACCAAGACTCAAC 1307  
Db 1264 GTACAGATTGGTGGTGGTACAGTGGCTCTCGGTACGGCGACGATGACAGCAC 1323  
QY 1308 TCTAAAGTATGAGTGTGCAGACGAGAAGAACTGGCACAGGAGTGCCTGGACATCGGA 1367  
Db 1324 ACTAAGGTGTGGGATGTGAAGGCCAGAGCTGGCCATGGACCTGCCCGCCACGCGGA 1383  
QY 1368 TGAGGTGTGGAGTGGACTGGCGCCGATGGCTCTAGAGTGGCTCTGTGGTGGCAAGGA 1427  
Db 1384 TGAGTATATGCTGTGACTGGAGTCCAGATGGCCAGAGTGGCAAGTGGTGGGAAGGA 1443  
QY 1428 CAAAGTTATAAGCTATGG 1446  
Db 1444 CAAATGCTCCGGATATGG 1462

## RESULT 13

AX405715  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
KEYWORDS  
FEATURES  
source  
CDS

AX405715  
Sequence 130 from Patent WO022660.  
AX405715  
AX405715.1 GI:21438869  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
Tang, Y.T., Liu, C., Zhou, P., Asundi, V., Zhang, J., Zhao, Q.A., Ren, F.,  
Xue, A.J., Yang, J., Wehrman, T. and Drmanac, R.T.  
Novel nucleic acids and polypeptides  
Patent: WO 022660-A 130 21-MAR-2002;  
HYSEQ, INC. (US)  
Location/Qualifiers  
1. 2638  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
82. 1539  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAD34822.1"  
/db\_xref="GI:21438870"  
/translation="MAAAVADEAVARDVQRLLVQFDEGQLGSPEDVPDITPDRLL  
QLVNALLAQEDPLEAFVYHDAETVSSLGKTLIESQAVETEKVLDIYQPAIFRVRA  
VTCTSLSEHSEAVISVAFSPKTYLASSGDTTVRFWDLSTETPFTCKGHRHWL  
SISWSPDKKLLASCKNGQILLWDPSTKQVGTILAGSKWITGLSWEPLHNPCEY  
VASSSKDGSRVIRWDTTAGRCIRLTGHTQSVTLRWGDDGLLYSASODRTIKVRAHD  
GVLCRTLOGHGHVNTMALSTDAIRLRTGAFEPAESVNPDIQSGLOELKRALSRYN  
LVRGQPERLVSGSDPFLTLPSPEDKKPLRMTGHQALINQVLPSPSRIVASNF  
DKSKLDGRTGKYLASLRHVAAYQIAMSADRLLYSGSSDSTLKWMDVRAQKLAM  
DLFGEDEVIAVDVSDGORVASGGKDKLIRWR"

## BASE COUNT

566 a 767 c 764 g 541 t

## ORIGIN

Query Match 30.7%; Score 477.4; DB 6; Length 2638;  
Best Local Similarity 59.8%; Pred. NO. 1.3e-108;  
Matches 837; Conservative 0; Mismatches 556; Indels 6; Gaps 2;

QY 54 GATACAGCGCGCTCTGTTCACGCGGAGGAGCGCGCCGCAATCGACCTGCGCGC 113  
Db 132 CTTAGTTCAGTTCCAGGATAGGGGGGCGAGCTGTGGGTTCCCGTTTCGACGTGCCCCGT 191  
QY 114 AGGAATCACTACCCAGCAATTTGGGACTGATTTGCAACGGCTGCTGAAAAACGAGGAGC 173  
Db 192 GGACATACCCCGGACAGGCTGCAGCTCGTGTCAACGGCTACTTGGCCCGCAGGAGATCC 251  
QY 174 CACTTCAATATTTTTCGTGGCGAGGATGAGATCAAGAAGAGCGCTCGAGGACACGTT 233  
Db 252 CCGCCACTGGCTTCTTTTTCACGATGCTGAGATGCTCTCTCACTTGGGAGAGGTT 311  
QY 234 GGATTTGGCTGAGTGGACACGCAAAACGCTGATGATTTGATATGACGACACAGCGGT 293  
Db 312 GGAGTCCCGAGTGGAGACAGAGAAGGCTCTAGACATCATCTACACGACACAGGCTAT 371  
QY 294 TTCAAAGTGGCGCCAGTGCAGAGTGCAGAGTTCATGCGCGGACACGCGGAGGCTGT 353  
Db 372 CTTCAAGTCCCGGCTGTGACTCGCTGCACACGCTCTTTGGAGGCTCACAGTGAGCGAGT 431  
QY 354 GGTTCGCTGAATTTTCAGCCCGGATGGTGTCTATCTCGCCAGTGGAGTGGCAGACACAC 413  
Db 432 CATTCTGTGGCTTCAGCCCTACCGGAAAGTACCTGGCCAGTGGCTCTGGAGACACAC 491  
QY 414 AGTCGATTTGGGATCTTAACACAGACACACGCACTTCACTTGCACAGTCTCATAGCA 473  
Db 492 CGTCGCTTCTGGGATCTCAGCAGACAGACACACATTTTCATGTCGAGGACACAGACA 551  
QY 474 GTGGTTCGTGGCTATCTCTGGGCTCCGGATGGCAACGGTTGGCCAGCGTGTGCAAGC 533  
Db 552 CTGGGCTCTTAGTATATCTGTCTCCAGATGGCAAGAGTGGCTCTCAGCTGCAAGAA 611  
QY 534 GGGCTCTATATCATCTGGGAGCGGCTCAGCAGAAAGGCGACCCCTTGATGG 593  
Db 612 TGGCAGATTTCTCTCTGGAGCCCAAGCAGAGGAGAGAGTGGCGAGGACCCCTCGCTGG 671  
QY 594 GCACAAGAAACACATCACTCGCTCGGTAACCGTATCATCGCATCCGATCCGAGTGCAG 653  
Db 672 CCACAGCAAGTGGATCACAGGCTGAGCTGGAGCCCTCCCATGCGAACCTGAGTGGCG 731  
QY 654 GAACTTGTCTCGCCAGTGGAGAGCGGGACTCCCGGATTTGGGACGTAAATTTGGCCA 713  
Db 732 CTATGGGCGCAGCAGCTCCAGGATGGAGTGTGGCGATCTGGGACACTGCGAGCGG 791  
QY 714 GTGCTTATGAATTCGGGACACACAAATGCTGTGACAGAGTGTGAGTGGGCTGAGC 773  
Db 792 CTGTAGCGCATCTCACCGGGCACACCCAGTCCGCTCCTCTCTCGGTTGGGAGGGGA 851  
QY 774 GGGCTTATTTATACATCTCAAGATCGCAGATGTAAGATGTGGCGAGCAGCTGATGG 833  
Db 852 CGGCTTCTCTACTCTGCTCCAGGACCGCACCATCAAGTCTGGAGAGTCTATGACGG 911  
QY 834 AATCTGTGGCGAGTCTCTGGCGAAGCTCACTGGTAAACACATTTGGCTGAGGAC 893  
Db 912 TGTCTGTGGCGAGCTCTGCAAGGCCAGCGCCACTGGGTGAACACCATTTGGCCCTCAGC 971  
QY 894 CGATTACGCTGCGGACTGTGCTTCCATCCGTTGAAGATGCTCTCAAGAGCC---A 950  
Db 972 TGACTATGCTCGCACTGGGGCTTTTGAACCTGTGAGGCGCTCAGTTAATCCCCAAGA 1031  
QY 951 CCTCAGTTTTCAGTCTGAGGATTTGAGGAACTCTGCTTTGAAGCGCTACACAGCGCTGTG 1010  
Db 1032 CCTCAGGATCTTTCAGGAGTGAAGGAGAGGGCTCTGAGCCGATACACCTCTGTCG 1091  
QY 1011 CCCTGAGGAGTGGAGTGTGCTGCTTCTCTGCGATGACACACCCCTCTATCTGTGG-- 1068  
Db 1092 GGGCAGGGTCCAGAGAGGCTGGTGTCTGGCTCCGAGACTTTCACCTATTCTCTGTGGTC 1151  
QY 1069 -CGGAACAACAGACAAGTGGCTTGGAGCGCATGACAGGCGACACAGAGCTGTGTCACGA 1127  
Db 1152 CCCAGCAGAGGACAAAAGCCCTCTCACTCGGATGACAGGACACCAAGCTCTCTCATCAACCA 1211  
QY 1128 TGTGAATATTCGCGGATGTAAGCTAAATTTGGTCTGCTTCTTTCATTTGACAGTCAAGTGG 1187

```

1212 GGTGCTCTTCTCTCGTACTCCCGCATCGTGGCTAGTGCCTCTTTGACAAATCATCA 1271
1188 TCTGTGGCGAGCGAGGATGGTCACTACATGCCACCTCCCGGGTCACTGTGCAGCGTGT 1247
1272 GCTGTGGGATGGCAGGAGCGGCAAGTACCTGGCTTCCTACGCGGCCAGTGGCTGCGGT 1331
1248 TTACACGGTTCCTTGGTCCCGGACTCCCGCTTATTTGTTTCCGGCAGCAAAAGACTCAAC 1307
1332 GTACAGATTGGTGGTTCAGCTGACAGTCGGCTCCTGGTCAGCGGCAGAGTGACGAC 1391
1308 TCTAAAGTATGAGTGTGCAGACGAAGAACTGGCAGCAGAGCTCCCTGGACATGCCGA 1367
1392 ACTGAAGTGTGGATGTGAAGCCCAAGCTGGCCATGGACTGCCCGGCCACGCGCA 1451
1368 TCAGGTGTTCGAGTGGACTGGCGCCGATGGCTCTAGAGTTGCTCTGTTGGTGGCAAGA 1427
1452 TGAGGTATATGCTGTGACTGGAGTCCAGATGGCCAGAGAGTGGCAAGTGGTGGGAAGA 1511
1428 CAAAGTTATAAGCTATGG 1446
1512 CAAATGCTCGGATATGG 1530

```

## RESULT 14

```

BC018399
LOCUS
DEFINITION Mus musculus, hypothetical protein similar to beta-transducin
family, clone MGC:25690 IMAGE:3491925, mRNA, complete cds.
ACCESSION BC018399
VERSION BC018399.1 GI:17390942
KEYWORDS MGC.
SOURCE house mouse.
ORGANISM Mus musculus.

```

```

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1784)
DIRECT SUBMISSION Strausberg, R.
TITLE Submitted (03-DEC-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgapbs@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hqsc.bcm.tmc.edu/cdna/
Contact: amg@bcm.tmc.edu
Gunaratne, P.H.; Garcia, A.M.; Lu, X.; Hulyk, S.W.; Hale, S.M.;
Yoon, V.S.; Kowis, C.R.; Lawrence, S.; Martin, R.G.; Muzny, D.M.;
Richards, S.; Gibbs, R.A.

```

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Series: IRAK Plate: 30 Row: b Column: 14.

## FEATURES source

```

Location/Qualifiers
1..1784
/organism="Mus musculus"
/db_xref="taxon:10090"
/map="CZCH II"
/clone="MGC:25690 IMAGE:3491925"
/tissue="Mammary tumor metastasized to lung. Tumor
arose spontaneously from a senescent normal mammary
(cloneal) outgrowth infected with the virus MMV."
/clone_lib="NCI CGAP_Lu29"
/lab_host="DH10B"
/note="vector: pCMV-SPORT6"

```

```

CDS
31..1488
/codon_start=1
/product="hypothetical protein similar to beta-transducin
family"
/protein_id="AAH18399.1"
/db_xref="GI:17390943"
/db_xref="LocusID:54475"
/translation="MAAAVVEAAAGDVORLLVQFQDEGOLLGSPDFVDPVDTTPDKL
OLVCNALLAQEPLPLAFYVHDAEIVSSLGKLTSESQVETEKIVDIIYQPAVFRVR
VTRCTSSLEGSEAVISVAPSTGKYLSAGSGDTTREFWDLSTPFTCTCKGHRHWL
SISPSFDGKLKASCKNGQIILLWDPTGLQVGRITLTGHSKWTGLSWPELHMPCRY
VASSSKDSGVKRWDTTAGRCERILDTGHTSVTCLRWGGGLLISASODRTIKYWRAD
LYLCRTGLQGHGHWNTMALSTYALRTAFPAEATVNAQDLQSLKELKERRASSRN
LVYRGGPRLVYSGDDFTLFLWPAEDKKPLARMTGHQALNQVLFSRIVASASF
DKSILKWDRTGYKYLASLRGHVAAVQIAWSDSRLVYSGSDSLKLVWVDAKQIAT
DLPFHADSVYAVDWSPDGQVRVAGGDKCLRIWR"
BASE COUNT 397 a 521 c 539 g 327 t
ORIGIN
Query Match 30.6%; Score 476.6; DB 10; Length 1784;
Best Local Similarity 59.2%; Pred. No. 1.9e-108;
Matches 850; Conservative 0; Mismatches 579; Indels 6; Gaps 2;
QY 18 GGAGACGGACACGGAGAGCGCCACATACATACAGGCGCCCTCGTTTACAC 77
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
45 GGTGAAGAGCGCGCGGACGCTGCAGCGGCTGCTGTACAGTTCAGGATGAGGG 104
QY 78 GGGGAGGAGCGCGCGGCGCCCACTCGACCTGCCGCGAGGATCACTACCCAGCAATGGG 137
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
105 CGGTCTGCTGCTGGGCTCCCGCTTCGACGTGCGCGTGACATCAGCCGGACAGCTGCA 164
QY 138 ACTGATTTCACACGCGCTGCTGAAAAACAGAGGAGCACTCCATATTTTGTTCGTGGG 197
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
165 GCTGTGTCAACGCGCTGCTGGCCAGAGGAGCGCCCTCGGCTTCTATGTCCTCA 224
QY 198 CGAGGATGATCACAAGAGCGCTGGAGGACAGTGGAGCTTGGCTGAGTGGACACCGA 257
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
225 CGATGCTGAGATTGTCTCCTCGGGGAGACGCTGGAGTCCAGTCTGTGGAGACAGA 284
QY 258 AAAGTGATGATATTTGTATGATGACCCACAGCGGCTTTCAAGTGGCGCCAGTGACAAG 317
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
285 GAAATGCTGATATCATCTACAGCCCAAGCTGTCTTCAAGGTCCTCGGGGTGAGCGCG 344
QY 318 ATGACAGAGTTCATCGCGGACACGCGGAGCTGTGTTCTGCTGAAATTTAGCCCGGA 377
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
345 GTGCACCAAGCTCATTTGGAGGCTCACAGCGAGCTGTCTTCTGTGGCTTCAGCCGAC 404
QY 378 TGGTGTCTATCTGCCAGTGGAGTGGCGACACACAGTGGATTTCTGGGATCTTACAC 437
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
405 AGAAAGTACTGCGCGAGTGGTTCGGAGACACACACGTGCTGCGCTTCTGGGATCTCAGC 464
QY 438 AGAGACACCGCACTTCACCTGCACAGTGCATAGCAGTGGGTTCTGCGGTATCTGGGC 497
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
465 TGAGACACCGCACTTCACATGCAAGGACACCGGCTGCTTCTAGCATATCTGTGTC 524
QY 498 TCCGAGTGGCAACAGGTTGGCCAGCGTTGCAAGCGGCTCTATATCTATCTGGGACCC 557
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
525 CCCAGATGGCAAGAACTGGCTCAGCTGCAAGANTGGCCAGATTTCTGCTGGGACC 584
QY 558 GGAGACGGGTACGACGAAGGCGACCTTGAAGTGGGCAACAAGAACATCACTGCTGCT 617
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
585 AAGCACGGGCTCAGGTGGCAGGACCCCTCACTTGGCCACAGCAATGATCAGAGCCT 644
QY 618 GCCTGGGAACCGTATCATCGGATCCGAGTCCAGGAAACTTGTTCGCCAGTGGA 677
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
645 GAGTGGGAGCCCTTCATCATGAACCCGCGAGTGGCTGCGGTAGTGGCCAGCAGCTCCAAGA 704
QY 678 CGGGGACTGCCGATTTGGGACGTAAATATGGGCCAGTGCCTTATGAACATTTGCCGACA 737
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
705 CGCAGTGTACGAGTCTGGGATACAAACCGCAGCGCTGTGAACGCACTCTCTCACGGGACA 764
QY 738 CACAAATGCTGTACAGCAGTGAAGTGGGTGGAGGGGCTTTATTTATACATCTCCAA 797
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```



Db 765 CACGACGTCAGTCACCTCGCTCGATGGGAGAGATGGCTTCTATATCTGCTTCCCA 824  
Qy 798 AGATCGCACAGTGAAGATGGCGACGACGATGGATCTTGTGCGGACGCTTCTG 857  
Db 825 GGACGAACCATCAAGCTTGGAGGCCCATGACGGGTGCTGTGCTGCTACTTCAAG 884  
Qy 858 CCAAGCTCACTGGGTAAACAACATTCGCTGAGCACCGGATACGTCCTCGCACCTGGTCC 917  
Db 885 TCATGGCCACTGGGTGAACACCATGCGCTTAGCACAGACTATGCCCTCGCACAGGGC 944  
Qy 918 ATTCCATCCGCTGAAGGATCGCTCAAGAGCC---ACCTCAGTTTTCGACGACTGAGAA 974  
Db 945 CTTTGAGCTGCTGAGCCACAGTGAATGCCAAGACCTCGAGGGTCTCTGAAGAGCT 1004  
Qy 975 CGAGAAATCTGCTTGAAGCGTACAGCGCTGTCCTGACGAGGTGAGTCTGCTGGT 1034  
Db 1005 GAAGGAGAGGCGATCGAGCGCTACAACTCGTCGGGGCCAAAGGCCACAGAGGCTGGT 1064  
Qy 1035 TTCTCTGTTTCGGATGACAACACCTCTATCTGTGG---CGGAACAACACAGAACTGGT 1091  
Db 1065 CTCTGGCTCAGAGACTTCACTTATCTGCTGTGCTCCAGAGAGGACAGAGCCCT 1124  
Qy 1092 TGAGCGCATGACAGGCGACAGAACGTTGTCACGATGTGAATATTCGCGGATGTAA 1151  
Db 1125 GCGCGGATGACAGGACACAGGCGCTCATCAACAGGCTCTCTCTCTGACTCGCG 1184  
Qy 1152 GCTAATTCGCTCTGCTTCAATTTGACAAGTCACTGCTGTGCGGAGCGAGGATGTCA 1211  
Db 1185 CATCGTGCCAGTGCCTCTTCGACAAGTCTATCAAGCTTGGGATGGCGGCGCAA 1244  
Qy 1212 GTACATGCGCACCTTCGCGGTCATGTCAGGCTGTTTACAGGTTGCTGTCGCGGA 1271  
Db 1245 GTACCTAGTCTTCCCTGCTGGCGCAGTGGCGCGTGTACAGATGCTGCTGCTGAG 1304  
Qy 1272 CTCGCCGTTGATTTTCGCGCAGCAAGACTCAACTTAAAGATGAGTGTGCGAGC 1331  
Db 1305 CAGCGGACTATTAGTCAGCGGACGAGTACAGCAGCTGAAGTGTGGATGTGAAGC 1364  
Qy 1332 GAAGAACTGGCACAGGAGCTGCTGCGGATGAGGATGCTGCGGATGCTGCTGCGG 1391  
Db 1365 CCAGAACTGGCACAGGAGCTGCTGCGGATGAGGATGCTGCTGCTGCTGAG 1424  
Qy 1392 GCGCGATGCTGAGATGCTGCTGCGGATGAGGATGCTGCTGCTGCTGCTGAG 1446  
Db 1425 TCCAGATGGCCAGAGTAGCAAGCGCGGGAAGCAAGTGCCTCGGATATGG 1479

## RESULT 15

AXI35807  
LOCUS AXI35807 1928 bp DNA linear PAT 29-MAY-2001  
DEFINITION Sequence 1 from Patent WO0132614.  
ACCESSION AXI35807  
VERSION AXI35807.1 GI:14272042  
KEYWORDS human.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1928)  
AUTHORS Utku,N.  
TITLE Novel genes tzap7/a, tzap7/b and tzap7 involved in t cell  
activation and uses thereof  
JOURNAL Patent: WO 0132614-A 1 10-MAY-2001;  
Utku, Nalan (DE)  
FEATURES  
Location/Qualifiers  
source  
1..1928  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
CDS  
92..11546  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAC39574.1"  
/db\_xref="GI:14272043"

/translation="MWIADAVARDVRLVQFODEGQQLGSPFDPVDPVITPDRLO  
LYCNALLAQEDPLSLAFFVHDABEIVSSLGKTLSEQVETEKVDLIIPTQAVFKVRV  
TRCTSSLEGHTAEVIVSFAFPTGKYLASGSDTTRVFWDLSTETPHPTSGHTHWVYL  
TAWSPDGKKLASGKNSQIFIMPSTGKIQKPLTGHSHKNITLWCPHLNAPESRYL  
ASASSGRVDRITWDTTGTAGTCRILTGTQSVTCRLWGGDGLLYSASODRTIKVYRAHDG  
VLCRTLOGHSHWNTMALSTDYALRTAFEPAEASVNPQDLOGSLQERALSRYNL  
VRGQSPERLVSQSDFTFLWSPAEKPKPLPMTHCHQALINQVLFSPDSIVASAFSD  
KSKLMDGRGKYLASLRHVAAYQIANSADSRLLVSGSSDSTLKVMDVKAQKLAND  
LPGHAEVYAVDVPDQGRVAGSGKDKLRLWRR"

BASE COUNT 432 a 540 g 550 g 406 t

Query Match 30.0%; Score 466.2; DB 6; Length 1928;  
Best Local Similarity 59.3%; Pred. No. 7.7e-106;  
Matches 830; Conservative 0; Mismatches 563; Indels 6; Gaps 2;

Qy 54 GATACAGGCGCGCTCTTTACAGGCGGAGAGCGCGCGCCCATCACTGCGCGG 113  
Db 139 GCTAGTCAGTTCCAGGATGAGGCGGCGAGCTGTGGGTCTCCCGCTTCAGCGTGG 198  
Qy 114 AGGAATCACTACCCAGCAATTTGGGACTGATTTGCAACGCGTGTGAAAAACAGGAAGC 173  
Db 199 GGACATCAACCGGACAGGCTGCGACTCGTTTGAACGCGCTACTGCCAGGAGATCC 258  
Qy 174 CACTCATATTTGTTTTCGTCGCGGAGGATGAGATCAAGAAGAGCTGCGGACAGCTT 233  
Db 259 CCGTCACTGCTTCTTTGTCACAGTCTGAGATCGTTTCTTCACTCGGAGAGAGCTT 318  
Qy 234 GGACTTGGCTGCTAGGACACCGAAACGCTGATGATTTGTGTATCAGCCACAGCGGT 293  
Db 319 GGAGTCCCGAGCAGTGGAGACAGAGAAGTCTCTAGACATCATTTTATCCACACAGCTG 378  
Qy 294 TTTCAAAGTGGCGCCAGTGCACAGATGCACAGTTCATGCCGCGGACACACCCGAGCTG 353  
Db 379 GTTCAAAGTCTGCTGTACACAGATGTACAGCTCATTTGGAGGACACACACGAGCTG 438  
Qy 354 GCTTTCGCTGATTTACGCGCGGATGGTCTCATCTCCAGTGGAGTGGGACACAC 413  
Db 439 TATTTACGATGCTTACGCGCCAACTGGAAAGTATTTGGCAAGTGTTCCTGGGACACTAC 498  
Qy 414 AGTCGATTTGGGATCTTAACACAGAGACACCGCATTTCACTGACAGCTCATGAAGCA 473  
Db 499 AGTCGCTTTTGGGATCTCAGCACAGAACTCCACATTTTACATCTAAAGGGCATACACA 558  
Qy 474 GTGGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 533  
Db 559 CTGGGTTCTCAGTATGCTTGTGCTCCAGATGGCAAAACTTGCCTCAGGATGTAAAA 618  
Qy 534 GGGCTCTATAATCATCTGGGACCCGGAGACGGGTACGAGAAAGGGCGACCTTGGTGG 593  
Db 619 TAGTCAGATCTTCAATTTGGGACCCAGACAGGGAAGCAGATTTGGCAAACTTAACAG 678  
Qy 594 GCACAAGAAACACATCAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 653  
Db 679 GCATCAAGTGGATTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 738  
Qy 654 GAAACTGCTTCCCGCAGTGGAGAGCGGAGTCCCGGATTTGGGACGTAAAAATGGGCA 713  
Db 739 ATACCTAGCCAGTGCCTCCAGCGCGCGCTGCGAGCTGCTGCGGATCTGGGACACAACT 798  
Qy 714 GTGCTTATGAACATTTGCGGACACACAAATGCTGTGACAGAGTGTGAGATGGGTGAGC 773  
Db 799 CTGTGAGCGCATCTCAGCGGACACACCCAGTCTGCTGCTGCTGCTGCTGCTGCTGCTG 858  
Qy 774 GGGCTTATTTATACATCTCCAAAGATCGCACAGTGAAGATGTGGGAGAGCTGTATGG 833  
Db 859 GCGGCTTCTTACTCTGCTCCAGGACCGCACCATTCAAGTCTGGAGAGCTCATGAGCG 918  
Qy 834 AATCTTGTGCGGACGCTTCTGCGCAAGCTCACTGGGTAAACAACATTCCTGCTGAGCAC 893  
Db 919 TGTGCTGCGGAGCTCTCAAGGCGCCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 978  
Qy 894 CGATTAGCTCTCTGCGGACTGGTTCATTCCATCCCGGTGAAGGATCGCTCCAGAGCC---A 950



```
Db      979  TGACTATGCCCTGGCAGCTGGGGCCTTTGAACCTGCTGAGGCGCTCAGTTAATCCCAAGA 1038
Qy      951  CCTCAGTTTGAGCAGCTGAGGAATTCAGGAATCTGCCCTTGAAGCGCTACAGGCCGCTGTG 1010
Db      1039  CCTCCAGGATCCTTGACAGAGTTGAAGGAGAGGGCTCTGAGCCGATACAACCTCGTGCG 1098
Qy      1011  CCCTGAGAGGTGGAGTGGCTGGTTTCCTGTTCGGATGACAACACCTCTATCTGTGG-- 1068
Db      1099  GGGCCAGGGTCCAGAGAGGTGGTGTCTGGCTCCGAGACTTCACCTTATCTCTGTGTCTC 1158
Qy      1069  -CGGAACAACCAAGCAAGTGGCTTGAGCGCATGACAGGGCACCAAGAAAGTGGTCAACGA 1127
Db      1159  CCCAGCAGAGGACAAAAGCCCTCTCACTCGGATGACAGGACACCAAGCTCTCATCAACCA 1218
Qy      1128  TGTGAATATTCGCCGATGTAAGCTAATTCGCTCTGCTTCAATTGACAAGTCAAGTGGC 1187
Db      1219  GGTGCTCTCTCTCTGACTCCGCGATCGTGGCTAGTGCCTCTTTGACAAGTCCATCAA 1278
Qy      1188  TGTGCGGAGCCAGCGATGGTCAGTACATGGCCACCTTCCGGGGTCATGTGCAGGCTGT 1247
Db      1279  GGTGTGGATGGCAGGACGGCAAGTACCTGGCTTCCCTACGGGGCCACGTGGCTGCCGT 1338
Qy      1248  TTACACGGTTGCCGTGGTCCGCGACTCCCGCTTGATTGTTTCCGGCAGCAAAAGACTCAAC 1307
Db      1339  GTACCAGATTGCGTGGTCACTGACAGTGGCTCCTGGTCAGCGGCGAGCAGTGACAGCAC 1398
Qy      1308  TCTAAAGTATGGAGTGTGCAGACGAGAACTGGCAGCAGGAGCTGCCCTGGACATGCCGA 1367
Db      1399  ACTGAAGGTGTGGGATGTGAAGGCCCAAGAGCTGGCCATGGACCTGCCGGCCACGCCGA 1458
Qy      1368  TGAGGTGTTCGGAGTGGACTGGGGCCCGATGGCTCTAGAGTTGCCCTCTGTGGTGGCAAGGA 1427
Db      1459  TGAGGTATATCTCTTGACTGGAGTCCAGATGGCCAGAGTGGCAAGTGGTGGGAAGGA 1518
Qy      1428  CAAAGTTATAAGCTATGG 1446
Db      1519  CAAATGCCCTCGGATATGG 1537
```

Search completed: June 2, 2003, 06:03:25  
Job time : 3995 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 2, 2003, 03:05:19 ; Search time 378 Seconds  
(without alignments)  
9264.173 Million cell updates/sec

Title: US-09-830-980-2

Perfect score: 1555

Sequence: 1 atttcccaaaatgcagga.....aaaaaaaaaaaaaaaaaaaa 1555

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq 101002.\*

1: /SID52/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SID52/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SID52/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SID52/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SID52/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SID52/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SID52/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SID52/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SID52/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SID52/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SID52/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SID52/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SID52/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SID52/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SID52/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SID52/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SID52/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SID52/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SID52/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SID52/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SID52/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SID52/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SID52/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SID52/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1555	100.0	1555	21	AA27739
2	1487.8	95.7	1992	23	ABL05339
3	1285.8	82.7	4164	23	ABL05338
4	601.2	38.4	6223	23	ABL05232
5	588.2	37.8	4141	23	ABL05214
6	478	30.7	1859	22	AAF85075
7	477.4	30.7	1853	22	AAH14166
8	477.4	30.7	1854	22	AAF58316
9	477.4	30.7	2638	24	ABN59719

10	466.2	30.0	1928	22	AAF85073
11	333.4	21.4	1726	21	AAC48725
12	331.8	21.3	1643	21	AAC39906
13	249.4	16.0	1170	22	AAF85074
14	176	11.3	399	21	AAH30571
15	164	10.5	547	22	AAH05727
16	155.8	10.0	1722	20	AAK30340
17	153.6	9.9	573	22	AAH98017
18	123.8	8.0	540	21	AAF07777
19	120.4	7.7	378	22	AAF67530
20	80.4	5.2	316	21	AAF11429
21	79.4	5.1	2984	22	AAK94739
22	79.4	5.1	3025	22	ABA06363
23	78.4	5.0	300	20	AAZ14360
24	78.4	5.0	764	22	AAK91829
25	78.4	5.0	764	22	AAK93229
26	78.4	5.0	1950	22	AAK53860
27	69	4.4	1925	22	AAK51876
28	68.4	4.4	279	23	ABV55135
29	68.4	4.4	2977	24	ABQ72680
30	68.2	4.4	1464	24	ABN59758
31	68.2	4.4	1751	22	AAF58329
32	68.2	4.4	2064	21	AAF22375
33	64	4.1	1591	23	ABL23519
34	62.6	4.0	462	23	ABV57002
35	62.2	4.0	559	23	ABV54541
36	62	4.0	1810	21	AAQ77522
37	61.8	4.0	2359	15	AAQ70726
38	61.4	3.9	821	22	AAK92334
39	61.4	3.9	821	22	AAK93787
40	61.4	3.9	1890	22	AAK94610
41	60.2	3.9	2359	17	AA42212
42	60.2	3.9	2359	18	AA779603
43	60.2	3.9	2826	23	ABL13261
44	60.2	3.9	3965	23	ABL11658
45	60.2	3.9	4826	23	ABL13260

#### ALIGNMENTS

RESULT 1  
AAA27739  
ID AAA27739 standard; cDNA; 1555 BP.

XX AAA27739;

AC AAA27739;

XX 29-AUG-2000 (first entry)

XX Drosophila Notchless (Nle) cDNA.

DE Notchless; Nle gene; Notch; signalling; neurodegenerative disease;

KW cancer; diagnosis; cytostatic; neuroprotective; gene therapy; ss.

XX Drosophila melanogaster.

XX Key Location/Qualifiers

FT CDS 13..1455

FT /tag= a

XX /trans\_except= (pos:859..861, aa:His)

XX WO200026364-A1.

PD 11-MAY-2000.

XX 03-NOV-1999; 99WO-IB01891.

XX 03-NOV-1998; 98GB-0024045.

XX (EUMO-) EURO MOLECULAR BIOLOGY LAB.

XX Cohen S, Bouwmeester A, Royet J;

Nucleotide sequenc  
Arabidopsis thalia  
Arabidopsis thalia  
Nucleotide sequenc  
Human colon cancer  
Human CDNA clone ( DNA encoding a hum  
Murine 7-transmem  
Fusarium venenatu  
Novel human polynu  
Aspergillus niger  
Human full-length  
Human CDNA SEQ ID  
Human gene express  
Human CDNA 5'-end  
Human CDNA clone r  
Human polynucleoti  
Human polynucleoti  
Human prostate exp  
Human MDDT encodin  
Novel human coding  
Human GTP-binding  
Human secreted pro  
Drosophila melanog  
Human prostate exp  
Human ORFX ORF3077  
TATA-binding prote  
Human CDNA 5'-end  
Human CDNA clone r  
Human full-length  
Drosophila TATA-bi  
TATA-binding prote  
Drosophila melanog  
Drosophila melanog  
Drosophila melanog

WPI: 2000-365613/31.  
P-PSDB; AAV79678.

Novel Notchless protein and nucleic acids encoding them useful for treating and preventing cancer and neurodegenerative diseases -

Claim 19; Page 44-45; 52pp; English.

The present sequence is that of cDNA encoding Notchless, a novel protein of Drosophila. The cDNA was isolated from an eye disc cDNA library by screening with a genomic DNA fragment that flanks the P element insertion site in Drosophila 1(2)AAK13714. Notchless was identified in a screen for dominant modifiers of a Notch mutant phenotype in the Drosophila wing. The mutant dominantly suppressed the wing notching phenotype of notchoid mutations, and the Notchless protein was shown to bind to the cytoplasmic domain of Notch. Notchless modified Notch signalling activity in a variety of CC Notch-dependent signalling process in both Drosophila and Xenopus embryos. The Notchless protein, and nucleic acids encoding it, can be used in methods for the diagnosis and therapy of certain diseases, particularly cancer and neurodegenerative diseases (claimed). A probe capable of screening for the Notchless gene (Nle), a cloning or expression vector comprising Notchless DNA, cDNA or RNA, a host cell, a transgenic animal, and a method for production of Notchless protein in the host cell are also claimed.

Sequence 1555 BP; 418 A; 373 C; 448 G; 316 T; 0 other;

Query Match 100.0%; Score 1555; DB 21; Length 1555;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1555; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

DR  WPI: 2000-365613/31.
XX  P-PSDB; AAV79678.
PT  Novel Notchless protein and nucleic acids encoding them useful for
PT  treating and preventing cancer and neurodegenerative diseases -
XX  Claim 19; Page 44-45; 52pp; English.
XX  The present sequence is that of cDNA encoding Notchless, a novel
CC  protein of Drosophila. The cDNA was isolated from an eye disc cDNA
CC  library by screening with a genomic DNA fragment that flanks the P
CC  element insertion site in Drosophila 1(2)AAK13714. Notchless was
CC  identified in a screen for dominant modifiers of a Notch mutant
CC  phenotype in the Drosophila wing. The mutant dominantly suppressed
CC  the wing notching phenotype of notchoid mutations, and the Notchless
CC  protein was shown to bind to the cytoplasmic domain of Notch.
CC  Notchless modified Notch signalling activity in a variety of
CC  CC Notch-dependent signalling process in both Drosophila and Xenopus
CC  embryos. The Notchless protein, and nucleic acids encoding it, can
CC  be used in methods for the diagnosis and therapy of certain diseases,
CC  particularly cancer and neurodegenerative diseases (claimed). A
CC  probe capable of screening for the Notchless gene (Nle), a cloning
CC  or expression vector comprising Notchless DNA, cDNA or RNA, a host
CC  cell, a transgenic animal, and a method for production of Notchless
CC  protein in the host cell are also claimed.
XX  Sequence 1555 BP; 418 A; 373 C; 448 G; 316 T; 0 other;

Query Match 100.0%; Score 1555; DB 21; Length 1555;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1555; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy  1  AATTCCTCCCAAAATGCAGGAGCGGACGAGGAGCAAGAGCCAGCCACATACGATACAG 60
Db  1  AATTCCTCCCAAAATGCAGGAGCGGACGAGGAGCAAGAGCCAGCCACATACGATACAG 60
Qy  61  GCGCGCTCTGTTACAGGCGGAGGAGGCGCGCGCCCAATCGACTGCGCGGAGGAATC 120
Db  61  GCGCGCTCTGTTACAGGCGGAGGAGGCGCGCGCCCAATCGACTGCGCGGAGGAATC 120
Qy  121  ACTACCCAGCAATTTGGACTGATTTGCAACGGCTGCTGTAACAGGAGGAGCACTCCA 180
Db  121  ACTACCCAGCAATTTGGACTGATTTGCAACGGCTGCTGTAACAGGAGGAGCACTCCA 180
Qy  181  TATTTGTTTTTCGTGGCGGAGGATGAGATCAAGAGAGCCCTGGAGGACAGCTTGGACTTG 240
Db  181  TATTTGTTTTTCGTGGCGGAGGATGAGATCAAGAGAGCCCTGGAGGACAGCTTGGACTTG 240
Qy  241  GCGTCAGTGGACACCCGAAACGATGATCGATATTTGATATGATCAGCCACAGCGGTTTCAA 300
Db  241  GCGTCAGTGGACACCCGAAACGATGATCGATATTTGATATGATCAGCCACAGCGGTTTCAA 300
Qy  301  GTGCGCCAGTGACAAGATGCACGAGTCCATGCGCGGAGACGCGCGAGGCTGTGTTTCG 360
Db  301  GTGCGCCAGTGACAAGATGCACGAGTCCATGCGCGGAGACGCGCGAGGCTGTGTTTCG 360
Qy  361  CTGAATTTACGCGCGGATGCTCATCTCGCGAGTGAAGTGGCGACACCACTGGCA 420
Db  361  CTGAATTTACGCGCGGATGCTCATCTCGCGAGTGAAGTGGCGACACCACTGGCA 420
Qy  421  TTGTGGGATCTTAACACAGAGACACCGCACTTCACCTGCACAGGTCAATAGCAGTGGTT 480
Db  421  TTGTGGGATCTTAACACAGAGACACCGCACTTCACCTGCACAGGTCAATAGCAGTGGTT 480
Qy  481  CTGTGGGTATCTGTGGCTCCGGATGCAACCGTTGGCCAGCGGTTCAAGCGGCTCT 540
Db  481  CTGTGGGTATCTGTGGCTCCGGATGCAACCGTTGGCCAGCGGTTCAAGCGGCTCT 540
Qy  541  ATAATCATCTGGGACCGGAGCGGTCAGCAAGAGGGCGACCCCTTGGTGGGACACAG 600
Db  541  ATAATCATCTGGGACCGGAGCGGTCAGCAAGAGGGCGACCCCTTGGTGGGACACAG 600
Qy  601  AATACATCACTGCTCGCTCGGTAACCGTATCATCGGATCCGGAGTGCAGGAACATT 660

```

RESULT 2

ABL05339

ID ABL05339 standard; cDNA; 1992 BP.

XX

AC ABL05339;

XX

DT 26-MAR-2002 (first entry)

XX

DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 10499.

```

601  AAACACATCACTGCTCGCTCGGAAACCGTATCATCGCATCGGAGTGCAGGAACATT 660
Qy  661  GCTTCGCCAGTGGAGAGCGGAGTCCGCGATTGGGACGTAAATATGGCCAGTGCCTT 720
Db  661  GCTTCGCCAGTGGAGAGCGGAGTCCGCGATTGGGACGTAAATATGGCCAGTGCCTT 720
Qy  721  ATGAACATTTGCCGGACACACAAATGCTGTGACAGCAGTGAATGGGGTGGAGCGGCTT 780
Db  721  ATGAACATTTGCCGGACACACAAATGCTGTGACAGCAGTGAATGGGGTGGAGCGGCTT 780
Qy  781  ATTATATACATCTCCCAAGATGCGACAGTGAAGATGTCGCGAGCAGCTGATGAATCTTG 840
Db  781  ATTATATACATCTCCCAAGATGCGACAGTGAAGATGTCGCGAGCAGCTGATGAATCTTG 840
Qy  841  TGGCGGACCTTCTCTGGCCAAAGCTCACTGGGTAAACAACTTGGCTGAGCAGCATTAC 900
Db  841  TGGCGGACCTTCTCTGGCCAAAGCTCACTGGGTAAACAACTTGGCTGAGCAGCATTAC 900
Qy  901  GTCTCGGCACCTGTGTCATTCATCCGCTGAAGATCGCTCCAGAGCAGCTCAGTTTG 960
Db  901  GTCTCGGCACCTGTGTCATTCATCCGCTGAAGATCGCTCCAGAGCAGCTCAGTTTG 960
Qy  961  AGCACTGAGGAATTCAGGAATCTGCCCTTGAAGCGTACCAGCGGCTGCGCTGAGCAG 1020
Db  961  AGCACTGAGGAATTCAGGAATCTGCCCTTGAAGCGTACCAGCGGCTGCGCTGAGCAG 1020
Qy  1021  GTGAGTCTGCTGTTTCTGTTTCGATGACAAACCCCTCTATCTGTGGCGGAACAACAG 1080
Db  1021  GTGAGTCTGCTGTTTCTGTTTCGATGACAAACCCCTCTATCTGTGGCGGAACAACAG 1080
Qy  1081  ACAAAGTGGTTCAGCGCATGACAGGCGACCAAGCGTCAACGATGCAAAATATTCG 1140
Db  1081  ACAAAGTGGTTCAGCGCATGACAGGCGACCAAGCGTCAACGATGCAAAATATTCG 1140
Qy  1141  CCGGATGTAAGCTAATTTGCTCTGCTTCAATTTGACAAGTCACTGCTGTGGCGAGCC 1200
Db  1141  CCGGATGTAAGCTAATTTGCTCTGCTTCAATTTGACAAGTCACTGCTGTGGCGAGCC 1200
Qy  1201  AGCGATGTCAGTACATGSCCAGCTTCCGGGTCATGTGAGGCTGTTTACACGTTGCC 1260
Db  1201  AGCGATGTCAGTACATGSCCAGCTTCCGGGTCATGTGAGGCTGTTTACACGTTGCC 1260
Qy  1261  TGTTCGCGGACTCCCGCTTGATTTTCGCGGAGCAAAAGCTCAACTCTAAAAGTATGG 1320
Db  1261  TGTTCGCGGACTCCCGCTTGATTTTCGCGGAGCAAAAGCTCAACTCTAAAAGTATGG 1320
Qy  1321  AGTGTGACAGCAAGAAACTGGCAGAGCTGCTGGACATGCGGATGAGGTCTTCGGA 1380
Db  1321  AGTGTGACAGCAAGAAACTGGCAGAGCTGCTGGACATGCGGATGAGGTCTTCGGA 1380
Qy  1381  GTGACTGGGCGCGCGATGGCTCTAGAGTTCCTCTGCTGTTGGCAGCAAAAGTATAAG 1440
Db  1381  GTGACTGGGCGCGCGATGGCTCTAGAGTTCCTCTGCTGTTGGCAGCAAAAGTATAAG 1440
Qy  1441  CTATGGGCTTATTAACAATCAATTAATTTGATACACGCTAAGAAATCTAGGAATAAG 1500
Db  1441  CTATGGGCTTATTAACAATCAATTAATTTGATACACGCTAAGAAATCTAGGAATAAG 1500
Qy  1501  TAAACGCTCTGATTAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1555
Db  1501  TAAACGCTCTGATTAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1555

```



```
Qy 1504 AACGTCCTGAGTAA 1518
Db 1519 AACGTCCTGAGTAA 1533

RESULT 3
ABL05338
ID ABL05338 standard; cDNA; 4164 BP.
XX
AC ABL05338;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 10496.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
XX
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR P-PSDB; ABB61235.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions
XX
PS Claim 1; SEQ ID NO 10496; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB5737-ABB72072).
CC
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 4164 BP; 947 A; 1030 C; 1101 G; 1086 T; 0 other;

Query Match 82.7%; Score 1285.8; DB 23; Length 4164;
Best Local Similarity 89.2%; Pred. No. 1.7e-275;
Matches 1505; Conservative 0; Mismatches 7; Indels 175; Gaps 4;

Qy 7 CAATAATGACGAGACGACGAGCAAGAGCCACGCCACATACGATACAGGCGCGC 66
Db 1019 CAGAAATGCGAGGACGAGCAGGAGCAAGAGCCACGCCACATACGATACAGGCGCGC 1078
Qy 67 CTCGTTT---ACACGGGAGGAGCGCGCGCCCAATCGACCTGCGCGGAGGAATCACT 123
Db 1079 CTCGTTTTCGACACGGGAGGAGCGCGCGCCCAATCGACCTGCGCGGAGGAATCACT 1138
Qy 124 ACCAGCAATTTGGGACTGATTTGCAACGGCTGCTGTAATAACAGGAGGAGCACTCCATAT 183
Db 1139 ACCAGCAATTTGGGCTGATTTGCAACGGCTGCTGTAATAACAGGAGGAGCACTCCATAT 1198
Qy 184 TTGTTTTTCGTGGCGGAGGATGAGATCAAGAGAGAGCCCTGGAGGACACGTTGGACTTGGCG 243
```

```
Db 1199 TTGTTTTTCGTGGCGAGGATGAGATCAAGAGAGCGTGGAGGACACGTTGGACTTGGCG 1258
Qy 244 TCAGTGGACACCGAAACAGTATGATATGATATGATATGATATGATATGATATGATATGAT 303
Db 1259 TCAGTGGACACCGAAACAGTATGATATGATATGATATGATATGATATGATATGATATGAT 1318
Qy 304 CGCCCAATGACAAGATGACAGAGTCCATCGCCGAGACACGCCAGGCTGGTGTGCTG 363
Db 1319 CGCCCAATGACAAGATGACAGAGTCCATCGCCGAGACACGCCAGGCTGGTGTGCTG 1378
Qy 364 AATTTACGCCGGATGGTCTCATCTCGCCAGTGAAGTGGGACACACAGTGGCATTTG 423
Db 1379 AATTTACGCCGGATGGTCTCATCTCGCCAGTGAAGTGGGACACACAGTGGCATTTG 1438
Qy 424 TGGGATCTTAACACAGAGACACCGCACTTCACTGCACAGTGTATAGAGTGGTTCG 483
Db 1439 TGGGATCTTAACACAGAGACACCGCACTTCACTGCACAGTGTATAGAGTGGTTCG 1498
Qy 484 TCGGTATCTCTGGGCTCCGGATGGCAAAACGGTTGGCCAGCGTTGCAAGCGGCTCTATA 543
Db 1499 TCGGTATCTCTGGGCTCCGGATGGCAAAACGGTTGGCCAGCGTTGCAAGCGGCTCTATA 1558
Qy 544 ATCATCTGSGACCGGAGACGGGTGACAGAGGAGGCGACCCCTTGAGTGGGCAACAAGAAA 603
Db 1559 ATCATCTGSGACCGGAGACGGGTGACAGAGGAGGCGACCCCTTGAGTGGGCAACAAGAAA 1618
Qy 604 CACATCAACTGCCCTCGCTGGGAAACCGTATCATCGGATCGGAGTGCAGGAACCTTGT 663
Db 1619 CACATCAACTGCCCTCGCTGGGAAACCGTATCATCGGATCGGAGTGCAGGAACCTTGT 1678
Qy 664 TCCGCGAGTGGAGACGGGACTGCGGGATTTGGGACGTFARAATTTGGGCCAGTGCCTTATG 723
Db 1679 TCCGCGAGTGGAGACGGGACTGCGGGATTTGGGACGTFARAATTTGGGCCAGTGCCTTATG 1738
Qy 724 AACATTCGCGGACACACAAATGCTGTGACAGAGTGTGAGTGGGAGTGGGAGCGGCTTAT 783
Db 1739 AACATTCGCGGACACACAAATGCTGTGACAGAGTGTGAGTGGGAGTGGGAGCGGCTTAT 1798
Qy 784 TATACATCTCTCAAAAGATGCACAGTGAAGATGTGGGAGCAGCTGATGAAATCTTTGTC 843
Db 1799 TATACATCTCTCAAAAGATGCACAGTGAAGATGTGGGAGCAGCTGATGAAATCTTTGTC 1858
Qy 844 CGGAGTCTCTGCGCAAGCTCAGTGGGTAAACAATTTGGCTGAGACACCGATACGTC 903
Db 1859 CGGAGTCTCTGCGCAAGCTCAGTGGGTAAACAATTTGGCTGAGACACCGATACGTC 1918
Qy 904 CTGCGCACTGGTCTCCATTCCTCCGTTGAAGGATCGCTCCAAGAGCCACCTCAGTTTGAGC 963
Db 1919 CTGCGCACTGGTCTCCATTCCTCCGTTGAAGGATCGCTCCAAGAGCCACCTCAGTTTGAGC 1978
Qy 964 A-----C 965
Db 1979 AGTAAGAAATTCATATCTCTATATATGCTGTATATCTCTAAAGAGATGTTTTPACAGC 2038
Qy 966 TGAGGAATTCAGGAATTCGCTTGAAGCGCTACAGGCGGTGTCCTGACGAGGTGGA 1025
Db 2039 TGAGGAATTCAGGAATTCGCTTGAAGCGCTACAGGCGGTGTCCTGACGAGGTGGA 2098
Qy 1026 GTCCGCTGTTTCTGTTGCGATGACACACCCCTCTATCTGTGCGGAGCAACACCAAGACAA 1085
Db 2099 GTCCGCTGTTTCTGTTGCGATGACACACCCCTCTATCTGTGCGGAGCAACACCAAGACAA 2158
Qy 1086 GTCCGTTGAGCGCATGACAGGCGACCAAGAGTGGTCAACAGTGTGAATATTCGCCGGA 1145
Db 2159 GTCCGTTGAGCGCATGACAGGCGACCAAGAGTGGTCAACAGTGTGAATATTCGCCGGA 2218
Qy 1146 TGTAAAGCTAATTTGGCTGCTTTCATTTGACAGTCACTGCTGCTGCGGAGCGACGGA 1205
Db 2219 TGTAAAGCTAATTTGGCTGCTTTCATTTGACAGTCACTGCTGCTGCGGAGCGACGGA 2278
Qy 1206 TGTGTCAGTACATGGCGCACTTCCGGGCTCATGTGCAAGGCTGTTTACAGGTTGCTGCTG 1265
```

Db	2279	TGTCAGTACATGCGCACCTTCCGGGTCATGTGCAGGCTGTTTACAGGTTGCTGGTC	2338
QY	1266	CGGGGACTCCGCTTGATGTTTCCGCGCAGCAAGACTCAACTCTAAA-----	1313
Db	2339	CGGGGACTCCGCTTGATGTTTCCGCGCAGCAAGACTCAACTCTAAAAGTTAGCAGCT	2398
QY	1314	-----AGTATGGAGTGTGCAGA	1330
Db	2399	AGTTTCAGTTTCTAGCGCTAGCTAATACATATCCTCTTGTCTTATGATGGAGTGTGCAGA	2458
QY	1331	CGAAGAACTGGCAGAGGAGTGCCTGCACATCGGATGAGGTGTTCCGAGTGCAGTGGG	1390
Db	2459	CGAAGAACTGGCAGAGGAGTGCCTGCACATCGGATGAGGTGTTCCGAGTGCAGTGGG	2518
QY	1391	CGCCCGATGGCTCTAGAGTTGCTCTGGTGGGAGCAAGATTAAGCT-----	1442
Db	2519	CGCCGATGGCTCTAGAGTTGCTCTGGTGGGAGCAAGATTAAGCTTAAGCTTT	2578
QY	1443	-----ATGGGCTTA	1451
Db	2579	GAGTAACGAGTTTCTATATAAATATACAGACTCTAATTGTATCTTTTCAGATGGCTTA	2638
QY	1452	TTACAAATCATTAACCTGTACACGGTAAGAAATACCTTAGGAAATAAAGTAAAGCTCT	1511
Db	2639	TTACAAATCATTAACCTGTACACGGTAAGAAATACCTTAGGAAATAAAGTAAAGCTCT	2698
QY	1512	GAGTAAA	1518
Db	2699	GAGTAAA	2705
RESULT 4			
ID	ABL05232	standard; cDNA; 6223 BP.	
XX	AC	ABL05232;	
XX	DT	26-MAR-2002 (first entry)	
XX	DE	Drosophila melanogaster expressed polynucleotide SEQ ID NO 10178.	
XX	KW	Drosophila; developmental biology; cell signalling; insecticide;	
XX	KW	pharmaceutical; gene; ss.	
XX	OS	Drosophila melanogaster.	
XX	PN	WO200171042-A2.	
XX	PD	27-SEP-2001.	
XX	PF	23-MAR-2001; 2001WO-US09231.	
XX	PR	23-MAR-2000; 2000US-191637P.	
XX	PR	11-JUL-2000; 2000US-0614150.	
XX	PA	(PEKE ) PE CORP NY.	
XX	PI	Venter JC, Adams M, Li PWD, Myers EW;	
XX	DR	WPI; 2001-656860/75.	
XX	DR	P-PSDB; ABB61129.	
PT	PT	New isolated nucleic acid detection reagent for detecting 1000 or more	
PT	PT	genes from Drosophila and for elucidating cell signalling and cell-cell	
PT	PT	interactions	
XX	XX	Claim 1; SEQ ID NO 10178; 21pp + Sequence Listing; English.	
XX	PS		
XX	CC	The invention relates to an isolated nucleic acid detection reagent	
XX	CC	capable of detecting 1000 or more genes from Drosophila. The invention is	
XX	CC	useful in developmental biology and in elucidating cell signalling and	
XX	CC	cell-cell interactions in higher eukaryotes for the development of	
XX	CC	insecticides, therapeutics and pharmaceutical drugs. The invention	
discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA			
sequences (ABL01840-ABL16175) and the encoded proteins			
(ABB57737-ABB72072).			
The sequence data for this patent did not form part of the printed			
specification, but was obtained in electronic format directly from WIPO			
at ftp.wipo.int/pub/published_pct_sequences.			
XX	Sequence	6223 BP; 1458 A; 1588 C; 1574 G; 1603 T; 0 other;	
Query Match 38.78; Score 601.2; DB 23; Length 6223;			
Best Local Similarity 99.08; Pred. No. 1.8e-125;			
Matches 616; Conservative 0; Mismatches 3; Indels 3; Gaps 1;			
QY	7	CAAAAATGCAAGAGACGACGAGCAAGAGCCACCCACATACAGTACAGGCGCGC	66
Db	5602	CAGAAAATGCAAGAGACGACGAGCAAGAGCCACCCACATACAGTACAGGCGCGC	
QY	67	CTCGTTT---ACAGGCGGAGGAGCGCGCGCCCAATCGACTTCCGCGCAGCAATCACT	123
Db	5662	CTCGTTTCCGAGACGCGGCGAGGAGCGCGCCCAATCGACTTCCGCGCAGCAATCACT	5721
QY	124	ACCCAGCAATTGGGACTGATTGCAACGCGCTGCTGAAAAACGAGGAGCCACTCCATAT	183
Db	5722	ACCCAGCAATTGGGCTCTGATTGCAACGCGCTGCTGAAAAACGAGGAGCCACTCCATAT	5781
QY	184	TTGTTTTTCGTGGGCGAGGTGAGATCAAGAGAGCTTGGAGGACAGCTTGGACTTGGCG	243
Db	5782	TTGTTTTTCGTGGGCGAGGTGAGATCAAGAGAGCTTGGAGGACAGCTTGGACTTGGCG	5841
QY	244	TCAGTGGACACCGAAAAAGCTGATGATATTTGTATFACGCCACAGCGCGGTTTCAAGTG	303
Db	5842	TCAGTGGACACCGAAAAAGCTGATGATATTTGTATFACGCCACAGCGCGGTTTCAAGTG	5901
QY	304	CGCCAGTGACAAGATGCACAGATTCCATGCCGGGACACGCCGAGGCTGTGTTTCGCTG	363
Db	5902	CGCCAGTGACAAGATGCACAGATTCCATGCCGGGACACGCCGAGGCTGTGTTTCGCTG	5961
QY	364	AATTCAGCCCGGATGGTGTCTATCTCCGAGTGGGAGTGGCGCACACACAGTGCAGTG	423
Db	5962	AATTCAGCCCGGATGGTGTCTATCTCCGAGTGGGAGTGGCGCACACACAGTGCAGTG	6021
QY	424	TGGGATCTTTACACAGAGACACCGCACTTCACCTGCACAGGTATACAGCTGGTCTG	483
Db	6022	TGGGATCTTTACACAGAGACACCGCACTTCACCTGCACAGGTATACAGCTGGTCTG	6081
QY	484	TGCGTATCTCTGGCTCCGATGGCAACGGTTGGCAGCGGTTGCAAAAGCGGCTCTATA	543
Db	6082	TGCGTATCTCTGGCTCCGATGGCAACGGTTGGCAGCGGTTGCAAAAGCGGCTCTATA	6141
QY	544	ATCATCTGGAGCCCGGAGACGCGGTTCAGCAGAGGGGGACCCCTTGGTGGGCACAGAA	603
Db	6142	ATCATCTGGAGCCCGGAGACGCGGTTCAGCAGAGGGGGACCCCTTGGTGGGCACAGAA	6201
QY	604	CACATCAACTGCTCTGCTGGG	625
Db	6202	CACATCAACTGCTCTGCTGGG	6223
RESULT 5			
ID	ABL05214	standard; cDNA; 4141 BP.	
XX	AC	ABL05214;	
XX	DT	26-MAR-2002 (first entry)	
XX	DE	Drosophila melanogaster expressed polynucleotide SEQ ID NO 10124.	
XX	KW	Drosophila; developmental biology; cell signalling; insecticide;	
XX	KW	pharmaceutical; gene; ss.	
XX	OS	Drosophila melanogaster.	
XX	XX		





XX Utku N;  
XX WPI; 2001-316410/33.  
XX P-PSDB; AAB68284.  
XX Novel polynucleotide encoding Tzap protein involved in T cell  
XX activation, useful for diagnosing and treating diseases involving T  
XX cell activation, for treating organ transplantation rejection,  
XX rheumatoid arthritis -  
XX Claim 1; Page 53-55; 68pp; English.  
XX The present sequence represents a cDNA clone encoding a Tzap protein.  
XX The Tzap gene is involved in T cell activation, and in the modulation  
XX of immune responses. Tzap polynucleotides and polypeptides are useful  
XX for diagnosing or treating acute and chronic diseases involving T cell  
XX activation and Th1 and Th2 immune response, for the treatment of acute  
XX and chronic rejection of allo- and xeno organ transplants and bone  
XX marrow transplantation, for the treatment of rheumatoid arthritis, lupus  
XX erythematosus, multiple sclerosis, encephalitis, vasculitis, diabetes  
XX mellitus, pancreatitis, gastritis, thyroiditis, for the treatment of  
XX malignant disorders of T, B or NK cells, for the treatment of asthma,  
XX lepromatosis, Helicobacter pylori associated gastritis or for the  
XX treatment of skin tumours, adrenal tumours or lung tumours, wound  
XX healing, growth disorders inflammatory and/or infectious diseases.  
XX Sequence 1859 BP; 392 A; 545 C; 566 G; 355 T; 1 other;  
Query Match 30.7%; Score 478; DB 22; Length 1859;  
Best Local Similarity 59.8%; Pred. NO. 6.8e-98;  
Matches 837; Conservative 0; Mismatches 556; Indels 6; Gaps 2;  
54 GATACAGCGCGCTGTTTACACGCGCGGAGGAGCGCGCCGCAATCGACCTGCGCGG 113  
Db 70 GCTAGTGCAGTTCACAGGATGAGGCGGCGAGCTGCTGGGTTCCCGCTTCGCGGT 129  
Qy 114 AGGAATCACTACCCAGCAATGGGACTGNTTTCAGCGCGCTGCTCAAAACGAGGAGC 173  
Db 130 GGACATACCCCGGACAGGCTGACGTGCTGTCAGCGCGCTACTGGCCGAGGATCC 189  
Qy 174 CACTCCATATTTTTCGTTGGGCGAGGATGAGATCAAGAAGAGCCTGGAGGACGTT 233  
Db 190 CTGCGCACTGGCTTCTTTCACAGCATGCTGAGATGCTCTCCTCACTGGGAGAGCTT 249  
Qy 234 GGACTTGGCGTCACTGGACACCGAAGAGCTGATCGATATTGTATCAGCCACAGCGGT 293  
Db 250 GGAGTCCCGAGCAGTGAGAGAGAGAGTCTTAGACATCATCTACCGCCACAGGCTAT 309  
Qy 294 TTTCAGTGGCGCCAGTGACAGATGACAGGATTCATCCCGGACAGCGCGGCTGT 353  
Db 310 CTTCAGATCCCGGCTGTGACTGCTGCACCGAGCTCTTGGAGGGTCAAGTGAGGAGT 369  
Qy 354 GGTTCGCTGAATTCAGCCGAGTGGTCTCATCTCGCCAGTGGAGTGGCGACACAC 413  
Db 370 CATTTCTGTGGCTTTCAGCCCTACGCGAAGTACTCTGGCCAGTGGCTCTGGAGACCC 429  
Qy 414 AGTGGATTTGGGATCTTAACACAGACACACCGCACTTCACCTGCACAGGTCTAAGCA 473  
Db 430 CGTGGCTTCTGGGATCTCAGCAGACAGACACACATTTTACATGCAAGGGACAGACA 489  
Qy 474 GTGGGTTCTGCGTATCTCCGCTCCGGATGGCAACGGTTGGCCAGCGGTTCGAAGC 533  
Db 490 CTGGGTCTTATGATATCTCTGGTCTCCAGATGGCAAGAGCTGGCTCAGGCTCGAGAA 549  
Qy 534 GGGCTCTATATCATCTGGGACCCGAGACGGGTGACAGAGAGGGCGACCTTGGTGG 593  
Db 550 TGGCCAGATTTCTCTCTGGGACCCAGACAGAGAGAGAGTGGGCGGACCCCTCGTGG 609  
Qy 594 GCACAGAAACACATCACTGCTCGCTGGGAGCCGCTATCATCGCGTGGGAGCGAG 653  
Db 610 CCACAGCAAGTGGATCACAGGCTGAGTGGAGAGCCCTCCATGCGAACCTTGAGTGGC 669

Qy 654 GAAACTTGCTTCCGCCAGTGGAGACGGGGACTCCGGATTTGGAGCTAAATTTGGCCA 713  
Db 670 CTATGTGGCCAGCAGCTCCAGAGTGGGAGTGGCAGTCTGGGACACACTGAGCGCCG 729  
Qy 714 GTGCTTATGAACATTTGCCGACACACAAATGCTGTGTGACAGAGTGAATGGGTGAGC 773  
Db 730 CTGTGAGCGCATCTCTACCGGGACACACCGTCTGCTCTCCGTGGGAGGGGA 789  
Qy 774 GGGCTTATTTATACATCTCCAAAGATGCGACAGTGAAGATGTGCGGAGAGAGTGTGG 833  
Db 790 CGGCTTCTCTACTGCTGCCCTCCAGGACCGCACCATCAAGTCTGGAGAGCTCATGCG 849  
Qy 834 AATCTTGTGGCGGACGTTCTCTGCGCAAGCTCACTGGTAAACACATTTGGCGCTGAC 893  
Db 850 TGTGCTGTGGCGGACGCTCGAAGGCCACGGCCTGTTGGTGAACACCTGCGCTCAGCAC 909  
Qy 894 CGATTACGCTCTGCGGACGCTGCTCCATTCCTCCGTTGAAGATGCTGCTCCAGAGCC--A 950  
Db 910 TGACTATGCGCTGCGGACGCTGTTGAACCTGCTGAGGCGCTCAGTTAATCCCAAGA 969  
Qy 951 CTTCAAGTTTGAAGCTGAGGAAATTCGAGGAATCTGCTTGAAGCGCTACCGAGCGGTGTG 1010  
Db 970 CTTCAAGATTCCTTTCAGGAGTTCGAGGAGGAGGCTCTGAGCGGATACCACTCGTGG 1029  
Qy 1011 CCTCAGAGGTGGAGTCTGCTGTTCTGTTGCGATGACACACCTCTATCTGTGG-- 1068  
Db 1030 GGGCCAGGTCAGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1089  
Qy 1069 -CGGACACACCAAGAGTGGTGGAGCGCATGACAGGCGCACAGAGTGGTGTACAGGA 1127  
Db 1090 CCGACAGAGAGACAAAAGCCCTCTCACTCGGATGACAGGACACCAAGCTCTCATCAACA 1149  
Qy 1128 TGTGAATATTCGCGGAGTAAAGCTAATTCGCTGCTGCTTCAATTTGACAGTCAAGTGG 1187  
Db 1150 GGTGNTCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1209  
Qy 1188 TGTGTGGCGAGCAGGAGTGTGCTAGTATGCGCCACCTCCGGGCTCATGTGCGAGCTGT 1247  
Db 1210 GCTGTGGAGTGGCAGGAGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1269  
Qy 1248 TTACAGGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1307  
Db 1270 GTACAGATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1329  
Qy 1308 TCTAAAGATGAGGAGTGTGCGAGACGAAAGAACTGGCAGAGGAGTCTGCTGAGACATGCG 1367  
Db 1330 ACTGAAGTGTGGAGTGTGAAGGCCAGAGCTGCGCATGCGCTGCGCGGACCGCGGA 1389  
Qy 1368 TGAGTGTGCGAGTGGAGTGGCGCGCGGCTGCTGAGATTTGCTGCTGCTGCTGCTGCTG 1427  
Db 1390 TGAGTATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1449  
Qy 1428 CAAAGTTAAGCTATGG 1446  
Db 1450 CAAATGCTCGGATATGG 1468

## RESULT 7

AAH14166

ID AAH14166 standard; cdna; 1853 BP.

AC AAH14166;

XX AC

XX DT 26-JUN-2001 (first entry)

XX DE

XX Human cdna sequence SEQ ID NO:11399.

XX KW

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX OS Homo sapiens.

XX XX

XX FN EPI074617-A2.

XX XX

PD 07-FEB-2001.  
XX 28-JUL-2000; 2000EP-0116126.  
XX 29-JUL-1999; 99JP-0248036.  
PR 27-AUG-1999; 99JP-0300253.  
PR 11-JAN-2000; 2000JP-0118776.  
PR 02-MAY-2000; 2000JP-0183767.  
PR 09-JUN-2000; 2000JP-0241899.  
XX (HELI-) HELIX RES INST.  
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Negai K, Otsuki T;  
XX WPI: 2001-318749/34.  
XX  
XX Primer sets for synthesizing polynucleotides, particularly the 5602  
PT full-length cDNAs defined in the specification, and for the detection  
PT and/or diagnosis of the abnormality of the proteins encoded by the  
PT full-length cDNAs -  
XX  
XX ClaIm 8; SEQ ID 11399; 2537pp + CD ROM; English.  
XX  
XX The present invention describes primer sets for synthesizing 5602  
CC full-length cDNAs defined in the specification. Where a primer set  
CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary  
CC to the complementary strand of a polynucleotide which comprises one of  
CC the 5602 nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to  
CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.  
XX  
SQ Sequence 1853 BP; 378 A; 549 C; 565 G; 361 T; 0 other;  
Query Match 30.7%; Score 477.4; DB 22; Length 1853;  
Best Local Similarity 59.8%; Pred. No. 9.2e-98;  
Matches 837; Conservative 0; Mismatches 556; Indels 6; Gaps 2;  
QY 54 GATACAGCGCGCTGTTACAGCGGAGGAGCGCGCGCGCAATCGACCTGCGCGG 113  
DB 80 GCTAGTCAGATCCAGAGATGAGCGCGGCGAGCTGCTGGGTTCCCGGTTCCGACGTGCCGT 139  
QY 114 AGGAATCAGTCACAGCAATGGGAGCTGATTGCAACGGCTGCTGAAAACAGGAGAGC 173  
DB 140 GGACATCACCCCGGACAGGCTCGAGCTCGTGTGCAAGCGCTACTTGGCCAGGAGATCC 199  
QY 174 CACTCCATATTTCTTTTCGTGGGAGGATGAGATCAGAGAGCGCTGGAGGACAGCTT 233  
DB 200 CTGCCACTGGCTTCTTTGTCACGATGCTGAGATCTGCTCTACTTGGGGAAGACGTT 259  
QY 234 GGACTTGGGCTCAGTGGACACCGAAAACGTGATCGATATGTTGATCAGGCACACGCGGT 293  
DB 260 GGAGTCCAGGAGTGGAGACAGAGAGGTCCTTAGACATCATCTACCGGAGGAGCTAT 319  
QY 294 TTTCAAGTCCGCCAGTGACAAGATGCACAGTTCATGCGCGGAGACACCGCGAGCTGT 353  
DB 320 CTTACAGATCCGGGCTGTGACTCGCTGCACACGCTCTCTTGGAGGAGTCCACAGTGAAGCACT 379

QY 354 GGTTTCGCTCAATTTCCACCCCGGATGGTCTCTATCTCCGACGTGGAAGTGGCAGACAC 413  
DB 380 CATTCTGTGGGCTTCAAGCCCTACGGGAAGTACCTTGGCCAGTCTCTGGAGACACAC 439  
QY 414 AGTCGATTTGGGATCTTAACACAGAGACACCGCACTTCACTTCCAGAGTCAATAGCA 473  
DB 440 CGTGGCTTCTGGGATCTCAGCAGACAGACACCAATTTCACTACATGCAAGGACACAGACA 499  
QY 474 GTGGTGTGTGGTATCCTCGGATGSCAAACGGTTGGCCAGCGGTTGCGGAGTGAAGC 533  
DB 500 CTGGTCTTAGTATATCTCTGCTCCAGATGCAAGAGCTGGCCCTCAGCTGCAAGAA 559  
QY 534 GGGCTCTATATCATCTCGGACCGGAGAGCGGTGAGCAGAGGGGAGCCCTTGAGTGG 593  
DB 560 TGGCCAGATTCCTCTGGGACCAAGCAGAGGAAGAGGTGGGAGAGCCCTCGCTGG 619  
QY 594 GCACAGAAACACATCAACTCGCTCGCTGGGAACCGTATCATCGCATCCGAGTGCAG 653  
DB 620 CCACAGCAAGTGGATCAGAGCCCTGAGCTGGGAGCCCTCCATCGCAACCTGAGTGGCG 679  
QY 654 GAAACTTGTCTCCGCCAGTGGAGAGCGGGAGTGGCGGATTTGGGACGTAAATTTGGGCCA 713  
DB 680 CTATGTGGCCAGCAGCTCCAAGGATGCGAGTGGCGATCTGGGACACAACTCGACGCCG 739  
QY 714 GTCCCTTATGAACATTTGGCGGACACACAAATGCTGTGACAGAGTGAAGTGGGTGGAGC 773  
DB 740 CTGTGAGCGCATCTCCACCGGGCACACCCAGTGGGTGACCTGTCTCCGGTGGGAGGGA 799  
QY 774 GGGCCTTATTTATACATCTCCAAAGATGCAAGTGAAGATGTGGGAGCAGCTGATG 833  
DB 800 CGGCTTCTACTCTGCTCCAGGACCGCACCACTCAAGTCTGGAGAGCTCATGCGG 859  
QY 834 AATCTGTGCGCGAGCTCTCTGGCAAGCTCACTGGGTAAACAACTTCGCTGAGAC 893  
DB 860 TGTGCTGTGCGGAGCTCTGAAGGCCACGCCACTGGGTGAACCACTGSCCTCAGCAC 919  
QY 894 CGATTAGCTCTGCGCACTGTCTCAATCCATCGGTGAGAGATCGCTCCAGAGACC--A 950  
DB 920 TGAATGCGCTGCGCACTGGGCGCTTGAACCTGCTGAGGCGCTCACTTANTCCCCAAGA 979  
QY 951 CCTCAGTTTGGAGCTAGAGAAATTTGAGGAATTCGCTTGAAGCGCTACACGCGCGTGTG 1010  
DB 980 CTTCCCAAGGATCTTTGCAAGGATTTGAAGGAGAGGGCTCTGAGCGCGATACAACTCTG 1039  
QY 1011 CCCTGACGAGGTGGAGTGCCTGCTGCTGCGATGACAAACACCTCTATCTGTGG-- 1068  
DB 1040 GGGCCAGGTTCCAGAGAGGCTGTGCTGGCTCCGACGACTTCACTTATCTGTGGTC 1099  
QY 1069 -CGGAACAACAGAAAGTGGTGGAGCGCATGACAGGACAGCAACGCTGTCAACGA 1127  
DB 1100 CCACAGAGGACAAAAGCCTCTCACTCGATGACAGGACACCAAGCTCTCATCAACCA 1159  
QY 1128 TGTGAATATTCGCGGATGTAAAGCTTAATTTGGCTGCTGCTTCAATTTGCAAGTCAAGTGG 1187  
DB 1160 GGTGCTCTTCTCTCTGACTCCCGCATCGGTGGCTAGTGCCTCTTTGACAAGTCCATCAA 1219  
QY 1188 TCTGTGGCAGCCAGATGGTCACTACATGCGCACTTCCGGGCTCATGTGCAAGGCTGT 1247  
DB 1220 GCTGTGGGATGCGAGGAGGCAAGTACCTTGGCTTCCCTACGCGGCCACGTGGCTGCCGT 1279  
QY 1248 TTACAGGTTGCTGCTCGCGGACTCCGCTTGAATTTTTCGCGAGCAGCAAGCACTACAC 1307  
DB 1280 GTACAGATGCTGGTTCAGCTGACAGTCCGCTTCTGTCAGCGGAGAGTGGCCATCGAGCAC 1339  
QY 1308 TCTAAAGATGAGTGTGACAGAGAAACTGGCAGACAGGAGTGCCTTGGACATCGGGA 1367  
DB 1340 ACTGAAGTGTGGGATGTGAAGGCCCAAGAGTGGCCATGACCTTGCCTGCCGCGCACGCGGA 1399  
QY 1368 TGAGGTGTTCGGAGTGGAGTGGGCGCGGATGGCTGTAGAGTTCCTCTGTGGGCAAGA 1427  
DB 1400 TGAGGTATATGCTGTGAGTGGAGTCCAGATGGCCAGAGAGTGGCAAGTGTGGGAAGA 1459  
QY 1428 CAAAGTTATAAGCTATGG 1446

```

Db      1460 CAAATGCCCTCCGGATATGG 1478
      |||| | | ||||
RESULT 8
AAF58316
ID      AAF58316 standard; cDNA; 1854 BP.
XX
AC      AAF58316;
XX
DT      19-APR-2001 (first entry)
XX
DE      Human GTP-binding associated protein #16 coding sequence.
XX
KW      Human; guanosine triphosphate binding associated protein; GTP; GBAF;
KW      inflammation; AIDS; Addison's disease; anaemia; arteriosclerosis; asthma;
KW      autoimmune disorder; hepatitis; multiple sclerosis; cancer; diabetes;
KW      osteoporosis; psoriasis; ss.
XX
OS      Homo sapiens.
XX
WO200105970-A2.
XX
PD      25-JAN-2001.
XX
PF      19-JUL-2000; 2000WO-US19698.
XX
PR      19-JUL-1999; 99US-0144595.
PR      23-AUG-1999; 99US-0150460.
PR      15-OCT-1999; 99US-0159849.
XX
PA      (INCY-) INCYTE GENOMICS INC.
XX
PI      Yue H, Tang YT, Bandman O, Hillman JL, Lal P, Au-Young J;
PI      Reddy R, Yang J, Baughn MR, Lu DAM, Azimzai Y, Patterson C;
XX
WPI: 2001-091972/10.
P-PSDB; AAB68516.
XX
PT      New guanosine triphosphate-binding associated proteins (GBAP) and their
PT      encoding nucleic acids, useful for treating and/or diagnosing diseases
PT      associated with GBAP expression, such as cancer, diabetes and asthma -
XX
PS      Claim 5; Page 197; 233pp; English.
XX
CC      The present invention relates to novel human guanosine triphosphate
CC      (GTP)-binding associated proteins (GBAPs; AAB68501-AAB68566) and their
CC      coding sequences (AAF58301-AAF58366). The proteins and coding sequences
CC      of the present invention are useful for treating a variety of disorders
CC      including inflammation, AIDS, Addison's disease, anaemia,
CC      arteriosclerosis, asthma, autoimmune disorders, grave's disease,
CC      hepatitis, multiple sclerosis, cancer, diabetes, osteoporosis and
CC      psoriasis.
XX
SQ      Sequence 1854 BP; 389 A; 548 C; 563 G; 354 T; 0 other;

Query Match      30.7%; Score 477.4; DB 22; Length 1854;
Best Local Similarity 59.8%; Pred. No. 9.2e-98;
Matches 837; Conservative 0; Mismatches 556; Indels 6; Gaps 2;

QY      54 GATACAGCGCGCTCGTTTACACGGCGGAGGAGCGCGCCCAATCGACCTCCGCGC 113
      |||| | | ||||
Db      69 GCTAGTGCAGTTCAGGATGAGGCGGCGCAGCTGCTGGGTTCCCGTTCGAGTGCCCGT 128
      |||| | | ||||
QY      114 AGGATCACTACCCAGCAATTTGGGACTGATTTGAACGCGCTCTCTGAAAACGAGGAAGC 173
      |||| | | ||||
Db      129 GGACATACCCCGGACAGGCTGCAGCTGTGTGCAACGCGCTACTGGCCAGAGGATCC 188
      |||| | | ||||
QY      174 CACTCCATATTTGTTTTCTGTGGCGGAGGATGAGATCAAGAAGAGCGCTGGAGGACAGTT 233
      |||| | | ||||
Db      189 CTGCGCACTGGCTTTCTTTGTCACGATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 248
      |||| | | ||||
QY      234 GGACTTGGGTCAGTGGACACCCGAAAACGTGATGATATTGTTGATATCACCACACAGCGGT 293
      |||| | | ||||

```

```

Db      249 GGAGTCCAGGCAGTGGAGACAGAGAAGGTCTTAGACATCATCTACCAGCCACAGGCTAT 308
      |||| | | ||||
QY      294 TTTCAAAGTCCGCCAGTGCACAAGATGCACAGTTCATCGCGGACACAGCCGAGGCTGT 353
      |||| | | ||||
Db      309 CTTCAGAGTCCGGGCTGTGACTCGCTGCACACAGCTCTTTGGAGGGTACAGTGAAGCAGT 368
      |||| | | ||||
QY      354 GGTTCGCTGAATTTTCAGCCCGGATGGTGTCTCATCTCGCCAGTGAAGTGGCGACACACCAC 413
      |||| | | ||||
Db      369 CATTTCTGTGGCCTTCAGCCCTACGGGAAAGTACCTGGCCAGTCTCTGGAGACACAC 428
      |||| | | ||||
QY      414 AGTCGATTTGGGATCTTAACACAGACACCCGCACTTCACTGCACAGTGCATGAAGA 473
      |||| | | ||||
Db      429 CGTCGGCTTCTGGGATCTCAGCACAGACACACACATTTACATCAAGGACACAGACA 488
      |||| | | ||||
QY      474 GTGGTCTCTCGGTATCTCGGGTCCGGATGGCAACGTTGGCGACGCTTGCACAAAGC 533
      |||| | | ||||
Db      489 CTGGGCTCTTAGTATATCTCTGCTCTCCAGATGGCAAGAAGCTGGCTCAGGCTGCAAGAA 548
      |||| | | ||||
QY      534 GGGCTCTATATCATCTGGGACCGGAGAGCGGTTCAGCAGAGAGGGGCGACCTTGAGTGG 593
      |||| | | ||||
Db      549 TGGCCAGTTCTCTCTGGGCCCAAGCACAGGAGAGAGTGGCGAGACCTCTGCTG 608
      |||| | | ||||
QY      594 GCACAAGAAACATCACTGCTCGCTGGGAACGATATCATCCGATCCGAGTCCAGTCCAG 653
      |||| | | ||||
Db      609 CCACAGCAAGTGTATCAGGCTGTAGCTGGGAGCCCTCCATGCGAACCCTGAGTCCGCG 668
      |||| | | ||||
QY      654 GAAACTTCTCCGCCAGTGGAGAGCGGGACTCGCGGATTTGGGACGTAATAATTTGGGCCA 713
      |||| | | ||||
Db      669 CTATGTGGCCAGCAGCTCCAGGATGGCAGTGTGCGGATCTGGGACACAACTGCAGCGCC 728
      |||| | | ||||
QY      714 GTGCTTATGAACATTTGCCGGACACACAATTTGTGACAGCAGTGCAGATGGGTGGAGC 773
      |||| | | ||||
Db      729 CTGTGAGCGCATCTCTCACCGGCGACACCCAGTCCGCTCTCTCGGTGGGAGGGGA 788
      |||| | | ||||
QY      774 GGGCTTATTTATACATCTCCCAAGTCCGACAGTGAAGATGTGGCGAGCTGATG 833
      |||| | | ||||
Db      789 CGGGCTTCTCTCTGCTCTCCAGGACCCGACCATCAAGTCTGGAGAGGCTCATGACGG 848
      |||| | | ||||
QY      834 AATCTGTCCGCGGCTTCTCTGCGCAAGCTCACTGGGTAAACACATCTGCGTGCAGTGCAC 893
      |||| | | ||||
Db      849 TGTGCTGTGCGGACCTCTCAAGGCCAGGCGCACTGGGTGAACACCATGGCCCTCAGCAC 908
      |||| | | ||||
QY      894 CGATTACGCTCTGCGCAGCTGCTCATTTCCATCCGGTGAAGGATCGCTCTCAAGAGCC - - - A 950
      |||| | | ||||
Db      909 TGACTATGCCCTCGCACCTGGGGCCTTTTGAACCTGCTGAGGCTCAGTGAATCCCAAGA 968
      |||| | | ||||
QY      951 CTTCAAGTTTCAGCACTGAGGAATTCAGGAATCTGCTTGAAGCGCTACACAGGCGTGTG 1010
      |||| | | ||||
Db      969 CTTCAAGGATCTTCGACGAGTTGAAGGAGGGGCTCTGAGCCGATACACCTCTGCTGCG 1028
      |||| | | ||||
QY      1011 CCCTGACAGGCTGAGTCCGCTGCTTCTGTTGGATGACACACCTCTATCTGTTGG - - 1068
      |||| | | ||||
Db      1029 GGGCCAGGGTCCAGAGAGGCTGTGTCTGCGTCCGAGACTTCACTTATTTCTCTGTTGTC 1088
      |||| | | ||||
QY      1069 -CGGAACAACAGAACAGTGGTTGAGCGCATGACAGGCGACAGAGACGTTGGTCAACGA 1127
      |||| | | ||||
Db      1089 CCCAGCAGAGGACAAAAGCCCTCTCACTCGGATGACAGGACACCAAGCTCTCATCAACCA 1148
      |||| | | ||||
QY      1128 TGTGAATATTTCCCGGATGTAAGAGCTAATTTGGCTGTGCTTCAATTTGACAGTCAAGTGC 1187
      |||| | | ||||
Db      1149 GGTGCTCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1208
      |||| | | ||||
QY      1188 TCTGTGGGAGCCAGGATGCTAGTACATGGCCACCTTCGCGGCTCATGTGCGAGTGT 1247
      |||| | | ||||
Db      1209 GCTGTGGGATGGCAGAGCGGCAAGTACCTTGGCTTCTTCCGCGGCGACAGTGGCTGCGGT 1268
      |||| | | ||||
QY      1248 TTACACGTTGCTGTGCTCGCGACTCCCGCTTGATTTGTTTCGCGCAGCAAGACTCAAC 1307
      |||| | | ||||
Db      1269 GTACCAAGTTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1328
      |||| | | ||||
QY      1308 TCTAAAGTATGAGTGTGCACAGCAAGAACTGGCACAGGAGTGCCTTGACATGCGGA 1367
      |||| | | ||||

```

Db	1329	ACTGAAGTGTGGGATGTCAGAGCCGAGAGCTGGCCATGAGACCTGGCCCGCCACCCGGA	1308
Qy	1368	TCAGGTTCGGAGTGGGCTGGCGCCGATGCTCTAGAGTGGCTCTGGTGGCAAGGA	1427
Db	1389	TCAGGTATATGCTGTGACTGGAGTCCAGATGGCCAGAGAGTGGCAAGGA	1448
Qy	1428	CAAAAGTTATAAGCTATGG	1446
Db	1449	CAAAATGCTCCGGATATGG	1467
RESULT 9			
ID	ABN59719	standard; cDNA; 2638 BP.	
XX	ABN59719;		
AC	ABN59719;		
DT	28-JUN-2002	(first entry)	
XX	Novel human coding sequence SEQ ID NO: 130.		
DE	Human; antianaemic; vulnerary; antiinflammatory; immunomodulator;		
XX	antinfertility; cerebroprotective; cytosstatic; rheumatic; gene therapy;		
KW	neuroprotective; antiparkinsonian; protein therapy; EST;		
KW	expressed sequence tag; gene; ss.		
XX	Homo sapiens.		
OS			
XX	WO200222660-A2.		
PN			
XX	21-MAR-2002.		
PD			
XX	10-SEP-2001; 2001WO-US26015.		
PF			
XX	11-SEP-2000; 2000US-0659671.		
PR			
XX	(HYSE-) HYSEQ INC.		
PA			
XX	Tang YN, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;		
PI	Xue AJ, Yang Y, Wehrman T, Drmanac RT;		
XX	WPI: 2002-292408/33.		
DR	P-PSDB; ABB97306.		
XX	An isolated polynucleotide for treating diseases associated with its		
PT	encoded polypeptide such as cancer and multiple sclerosis -		
XX	Claim 1; SEQ ID NO 130; 509pp; English.		
PS			
XX	The present invention provides the protein and coding sequences of 444		
CC	novel human proteins. These were isolated from expressed sequences tags		
CC	(ESTs). They can be used to stimulate cell growth, to regulate		
CC	haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth		
CC	e.g. in burn treatment, to regulate the immune system e.g. to treat		
CC	multiple sclerosis, to regulate activin or inhibin e.g. to treat		
CC	infertility, to regulate haemostasis or thrombolysis e.g. to treat		
CC	stroke and cancer, to screen for drugs, to treat inflammatory conditions		
CC	e.g. rheumatoid arthritis, and to treat nervous system disorders e.g.		
CC	Parkinson's disease. The present sequence is a coding sequence of the		
CC	invention.		
XX			
SQ	Sequence 2638 BP; 566 A; 767 C; 764 G; 541 T; 0 other;		
Query Match 30.7%; Score 477.4; DB 24; Length 2638;			
Best Local Similarity 59.8%; Pred. No. 1e-97;			
Matches 837; Conservative 0; Mismatches 556; Indels 6; Gaps 2;			
Qy	54	GATACAGGGCGCCTCGTTTACAGGGCGGAGGAAGCGCGCCCAATCGACCTGCCGCG	113
Db	132	GCTAGTGCAGTCCAGGATGAGGCGGCGAGCTGCTGGGTTCCCGCTTCGACGTGCCGT	191
Qy	114	AGGAATCACTACCAGCAATTTGGACTGATTTGCAACGCGCTGCTGAATAACGAGGAGC	173







```
QY 849 GTTCTCTGGCCAGCTCACTGGGTAAACAACATTCGGCTGAGCCAGCATTAAGTCTCTGGC 908
   || || || || || || || || || || || || || || || || || || || || || ||
Db 883 ATTGAAGGGCATGGGCATTTGGATTAACTCCCTTGGTTGAGCAGACAATATGTTCTTCG 942
   || || || || || || || || || || || || || || || || || || || || || ||
QY 909 CACTGGTCCATTCCATCCGGTGAAGGATCGCTCCAGAGCCACCTCAGTTGTGAGCAGTGA 968
   || || || || || || || || || || || || || || || || || || || || || ||
Db 943 AACAGGAGCTTTTGACC-----ACACTGGAGACAATATCTCCAAATGAGAAGAAGCAA 997
   || || || || || || || || || || || || || || || || || || || || || ||
QY 969 GGAATTCAGGAATCTGCTTGAAGCCCTACCAAGCCGTGTGCCCTGACGAGTGGATGC 1028
   || || || || || || || || || || || || || || || || || || || || || ||
Db 998 AAGCGCTCTGAAGATACAAACAAGGGGATTCCTCCCTGAAGATAGTGTCTCAGGT 1057
   || || || || || || || || || || || || || || || || || || || || || ||
QY 1029 GCTGGTTCTGTTCCGATGACAAACCCCTCTATCTGTGGCGGAACCAACCAAGTGG 1088
   || || || || || || || || || || || || || || || || || || || || || ||
Db 1058 TCTGATGATTTCACTATGTTCTCTTTGGGAACCATCTGTAGCAACAACCTAAA----- 1111
   || || || || || || || || || || || || || || || || || || || || || ||
QY 1089 CGTTGAGCGCATGACAGGCGCACCAAGACGTGGTCAACGATGTGAATATTCGCGCGATGT 1148
   || || || || || || || || || || || || || || || || || || || || || ||
Db 1112 ----AAGCGTTAAACCGGTATCAACAGCTTGTAAATCATGTCTATTTCTCACTGTATGG 1167
   || || || || || || || || || || || || || || || || || || || || || ||
QY 1149 AAAGCTAATTCGCTCTGCTTCATTGACAAGTCAGTGGCTGTGGCGAGCCAGCGATGG 1208
   || || || || || || || || || || || || || || || || || || || || || ||
Db 1168 GAATGGAATGCAAGTGCATGTTTCGATTAATCAGTTAGGCTATGGATGGTATCACAGG 1227
   || || || || || || || || || || || || || || || || || || || || || ||
QY 1209 TCAGTACATGGCCACCTTCGCGGTCTATGTGCAGGCTGTTTACAGGTTGCCCTGTGTCGC 1268
   || || || || || || || || || || || || || || || || || || || || || ||
Db 1228 ACATTTGTTACAGTTTCCCGGGACATGTTGGACCTGTTTATCAGGTCAAGTTGTGTCGC 1287
   || || || || || || || || || || || || || || || || || || || || || ||
QY 1269 GGACTCCCGTGTGTTTCCGGCAGCAAGACTCAACTCTAAAGTATGGAGTGTGCA 1328
   || || || || || || || || || || || || || || || || || || || || || ||
Db 1288 AGACAGTGAATGCTTTTGAGTGGCAGTAAAGACTCTACTCTCAAGATATGGGAATTAG 1347
   || || || || || || || || || || || || || || || || || || || || || ||
QY 1329 GACGAAGAAGTGGCAGCAGGAGTGCCTGGACATCGGATGAGTGTTCGGAGTGGATGG 1388
   || || || || || || || || || || || || || || || || || || || || || ||
Db 1348 GACGAAGAAGTAAACAAGATCTTCTGTCATCGATGAGTGTTCGCGTGGATGG 1407
   || || || || || || || || || || || || || || || || || || || || || ||
QY 1389 GGCSCCGATGCTCTAGAGTTCCTGTGGTGGCAAGCAAGTATATAAAGTATGGC 1448
   || || || || || || || || || || || || || || || || || || || || || ||
Db 1408 GAGTCCAGATGGAGAGAAAGTAGTTTCTGGTGAAGATAGATGTGTGACGCTATGAA 1467
   || || || || || || || || || || || || || || || || || || || || || ||
QY 1449 TTATTACA 1457
   || || || || || || || || || || || || || || || || || || || || || ||
Db 1468 GGTTTAAA 1476
   || || || || || || || || || || || || || || || || || || || || || ||

RESULT 12
AAC39906
ID AAC39906 standard; DNA; 1643 BP.
XX AAC39906;
AC
XX
XX
XX 17-OCT-2000 (first entry)
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 26334.
XX
XX Hybridisation assay; genetic mapping; gene expression control;
XX protein identification; signal transduction pathway;
XX metabolic pathway; promoter; termination sequence; ss.
XX
XX Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX
99US-0127462.
99US-0128234.
99US-0128714.
99US-0129845.
99US-0130077.
99US-0130449.
99US-0130510.
99US-0130891.
99US-0131449.
99US-0132048.
99US-0132407.
99US-0132484.
99US-0132485.
99US-0132487.
99US-0132863.
99US-0134256.
99US-0134218.
99US-0134219.
99US-0134321.
99US-0134370.
99US-0134768.
99US-0134941.
99US-0135124.
99US-0135353.
99US-0135629.
99US-0136021.
99US-0136392.
99US-0136782.
99US-0137222.
99US-0137528.
99US-0137502.
99US-0137724.
99US-0138094.
99US-0138540.
99US-0138847.
99US-0139119.
99US-0139452.
99US-0139453.
99US-0139492.
99US-0139454.
99US-0139455.
99US-0139456.
99US-0139457.
99US-0139458.
99US-0139459.
99US-0139460.
99US-0139461.
99US-0139462.
99US-0139463.
99US-0139750.
99US-0139817.
99US-0139817.
99US-0139899.
99US-0140353.
99US-0140354.
99US-0140695.
99US-0140823.
99US-0140991.
99US-0141287.
99US-0141842.
99US-0142154.
99US-0142390.
99US-0142390.
99US-0142803.
99US-0142920.
99US-0142977.
99US-0143542.
99US-0143624.
99US-0144005.
99US-0144085.
99US-0144086.
99US-0144325.
```





QY	729	TGCGGACACACAATGCTGTGCACAGCTGAGATGGGTGGAGCGGCGCTATTATTATAC	788
DB	730		
DB	763	CAGTGGGACACACTGCTGTGACCTTGTGTGCAATTTGGGGGAGATGGAATATTATTATAC	832
QY	789	ATCTCTCCAAAGATGSCACAGTGAAGATGTGGGAGCAGCTGATGGAATCTTGTGCCGAC	848
DB	823	AGTGTGCAAGATTTATGAGATTAAAGATGTGGGAGACTACTCAGGGGAAGCTATTATTCTGGA	882
QY	849	GTTCTCTGGCCAGCTCACTGGGTAAACAACATTCGCTGAGCAGCAGATAGCTGCTCGC	908
DB	883	ATTGAAGGGGATGGGCATTGGAATTAACTCCCTTCGCTTGAGCACAGAATATGTTCTTCG	942
QY	909	CACTGTGTCATTCCATTCGCGTGAAGGATCGCTCAAAGAGCCACCTCAGTTTCAGCAGCTGA	968
DB	943	AACAGGAGCTTTTGACC-----ACACTGGAAGACAAATATCTCCAAATGAAGAAAGCAA	997
QY	969	GGAAATGTCAGGAATCTGCTTTGAAGCGCTACCAAGCCGCTGTGGCCCTGAGAGGTGGATC	1028
DB	998	AAGCGCTCGAAGATACAACAACAAGGGGATTCCTCTGAAGATTAGTCTCAGT	1057
QY	1029	CTGTGTTCTCTGTCGGATGACACACCCCTCTACTCTGTGGGGAACAACGACAGCAAGTG	1088
DB	1058	TCGTGATGTTTCACTATGTTCCCTTGGGAACCATCTGTTAGCAACACACTAAA-----	1111
QY	1089	CGTTGAGCCGATGACAGGGACACAGAACGTGTCAACGATGTGAATATATCCCGGATGT	1148
DB	1112	----AAGCGTTAACCAGGTCATCAACAGCTTGAATCAATCTATTTCTCACTCTGATGG	1167
QY	1149	AAGCTAAATGTGCGTCTGCTTCATTTGACAAGTCAGTGGCTGTGGGAGCCAGCGATGG	1208
DB	1168	GAATAGATGTCGAATGCATCGTTTCGATAAATCACTAGTTATGGAATGGTATCACAGG	1227
QY	1209	TCAGTACATGCCACCTTCCGGGGTCATCTGCAGGCTGTTTACACGAGTTCCTGTGTCGC	1268
DB	1228	ACAAATTGTTACAGATTTTCCGGGGCATCTTGGACCTGTTTATCAGGTCAGTTGTCGCG	1287
QY	1269	GGACTCCCGCTTGATTGTTTCCGCGACGAAAGACACTCAACTCTAAAAGTATGAGAGTGC	1328
DB	1288	AGACAGTAGATTGCTTTTGTAGTGCGAGTAAAGACTCTACTCTCAAGATATGGAAATTAG	1347
QY	1329	GACGAAGAAGCTGGCACAGGAGCTGCCTGGACATCGGATGAGGTGTTCCGGAGTGGACTG	1388
DB	1348	GACGAAAAAGTTTAAACAAGATCTTCTGTCATCTGATGAGGTGTTTTCGGGTGGAATG	1407
QY	1389	GGCCGCCGATGCGCTAGAGTGTGCTGTGGTGGCAAGCAAAAGTTATATAAGCTATGGC	1448
DB	1408	GAGTCCAGATGAGAGAAAGTAGTGTCTGGTGGTAAAGATAGAGTGTTCGAAGCTATGAA	1467
QY	1449	TTATTAAACA	1457
DB	1468	GGGTTAAAA	1476
RESULT	13		
AA85074			
ID	AA85074	standard; DNA; 1170 BP.	
AC	AA85074;		
XX			
DT	09-JUL-2001	(first entry)	
DE			
XX	Nucleotide sequence of human Tzap gene cDNA clone Tzap7/A.		
KW	Tzap: T cell activation; immune response; transplant rejection;		
KW	bone marrow transplantation; rheumatoid arthritis; lupus erythematosus;		
KW	multiple sclerosis; encephalitis; vasculitis; diabetes mellitus;		
KW	pancreatitis; gastritis; thyroiditis; malignant disorder; asthma;		
KW	lepraematosus; gastritis; skin tumour; adrenal tumour; lung tumour;		
KW	wound healing; growth disorder; inflammatory disease; infectious disease;		
OS	ss.		
XX			
OS	Homo sapiens.		
XX			

Db 361 TGGGTCTGGCTCCGACGACTTACCTTATCTCTGTGGTCCCGACGAGGACAAAAGC 420  
 Qy 1088 CGGTGAGCGATGACGAGCGACCAAGAGTGTCAACGATGTGAAATATTCGCGGATG 1147  
 Db 421 CTCACACTCGGATGACGAGCACCAAGCTCTCATCAACGAGTGCTCTTCTCTCTGACT 480  
 Qy 1148 TAAAGCTAATGCGTGTCTTCATTTGACAACTGCTGCTGCTGCGGAGCGAGCGATG 1207  
 Db 481 CCCGATCGTGGCTAGTGCTCTTTCGACAACTGCTGCTGCTGCGGAGCGAGCGG 540  
 Qy 1208 GTACGATACATGCGACCTTCGCGGCTCATGTGCGAGCTGTTTACACGGTGTGCTGCTCG 1267  
 Db 541 GCAAGTACCTGGCTCCCTACGCGGCCACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 600  
 Qy 1268 CGGACTCCGCTGTATGTTTCCGCGACGACAACTCACTCTAAAGTATGAGTGTGC 1327  
 Db 601 CTGACAGTCCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 660  
 Qy 1328 AGCAGAACTGCGACGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1387  
 Db 661 AGCCCAAGAGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720  
 Qy 1388 GGGGCGCCGATGCTGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1446  
 Db 721 GGAATCCAGATGCCAGAGAGTGGCAAGTGGTGGGAGGACAAATGCTCCGCGATATGG 779  
 RESULT 14  
 AAH30571  
 ID AAH30571 standard; cDNA; 399 BP.  
 XX  
 AC AAH30571;  
 XX  
 DT 27-JUL-2001 (first entry)  
 XX  
 DE Human colon cancer cell line Km12L4-A cDNA library derived sequence #505.  
 XX  
 KW Human; diagnosis: colon cancer; cancer; malignant; chromosome mapping;  
 KW detection; colon cancer cell line Km12L4-A; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN W0200018916-A2.  
 XX  
 PD 06-APR-2000.  
 XX  
 PF 23-SEP-1999; 99WO-US22226.  
 XX  
 PR 28-SEP-1998; 98US-0102161.  
 PR 28-SEP-1998; 98US-0102180.  
 PR 29-SEP-1998; 98US-0102380.  
 PR 08-OCT-1998; 98US-0103815.  
 PR 27-OCT-1998; 98US-0105877.  
 XX  
 PA (CHIR ) CHIRON CORP.  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;  
 PI Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;  
 PI Lamson G, Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;  
 PI Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;  
 XX  
 DR WPI; 2000-293155/25.  
 XX  
 PT Polynucleotide library comprising 1079 defined sequences, useful in  
 PT the form of an array to detect cancer or susceptibility to cancer -  
 XX  
 PS Claim 1; Page 328-329; 502pp; English.  
 XX  
 CC The present invention describes a library of polynucleotides comprising  
 CC 1079 nucleotide sequences (given in AAH30067 to AAH31145). Also described  
 CC are: (1) an isolated polynucleotide (I) having at least 90% identity to

CC one of the 1079 sequences; (2) a recombinant host cell containing (1);  
 CC (3) an isolated polypeptide (II) encoded by (I); (4) an antibody that  
 CC specifically binds to (II); (5) a vector comprising (I); and (6) a method  
 CC of detecting differentially expressed genes correlated with a cancerous  
 CC state of a mammalian cell comprising detecting a gene product encoded by  
 CC 65 of the 1079 sequences given in the specification. The polynucleotides  
 CC are used to monitor patients having (or susceptible) to cancer to detect  
 CC potentially malignant events at a molecular level before they are  
 CC detectable at a gross morphological level. The polynucleotides are also  
 CC useful for monitoring the efficacy of various therapies and preventive  
 CC interventions. Polynucleotide probes based on the disclosed sequences  
 CC are useful for chromosome mapping and detection of transcription levels.  
 CC The 1079 polynucleotide sequences were derived from a human colon cancer  
 CC cell line Km12L4-A cDNA library.  
 XX  
 SQ Sequence 399 BP; 92 A; 120 C; 112 G; 75 T; 0 other;  
 Query Match 11.3%; Score 176; DB 21; Length 399;  
 Best Local Similarity 66.1%; Pred. No. 3.9e-30;  
 Matches 254; Conservative 0; Mismatches 130; Indels 0; Gaps 0;  
 Qy 266 TCGATATGTGTATCAGCCACAGCGGGTTTCAAGTGGCCAGTGACAGATGACAGA 325  
 Db 16 TAGACATCATCTACAGCCACAGGCTATCTCAGAGTCCGGCTGCTGACTCGCTGACCA 75  
 Qy 326 GTTCATGCGCGGACACCGCGAGGCTGTGGTTCGCTGAATTCAGCCCGGATGGTGCT 385  
 Db 76 GCTCCTTGGAGGGTCACAGTGAGGAGTCAATTTCTGTGGCTTCAGCCCTACGGAAAGT 135  
 Qy 386 ATCTCGGCAGTGGAGTGGCCACACAGTGGGATTTGGGATCTTAACACAGAGACAC 445  
 Db 136 ACCTGGCCAGTGGCTCTGGAGACACACCGTGGCTTCTGGGATCTCAGCAGACAGAC 195  
 Qy 446 CGCCTTCACCTGCACAGTCTAAGCAGTGGGTCTGTGCTATCTCTGGGCTCCGATG 505  
 Db 196 CACATTTTCACATCAAGGACACACAGACACTGGGCTCTTAGTATATCTGTCTCCAGATG 255  
 Qy 506 GCACACGTTGGCCACCGCTTCCAAAGCGGCTCTATATCATCTCTGGGACCGGACAGCG 565  
 Db 256 GCAGAGAGTGGCTCAGGCTGCAAGATGGCCAGATTTCTCTCTGGGACCCCAAGCAG 315  
 Qy 566 GTACAGCAGAGGGCGACCTTGTAGTGGGCACAGAAACACATCACTGCTGCTGCTGGG 625  
 Db 316 GGAAGCAGGTGGGACGAGGCCCTCGCTGGCCACAGCAAGTGGATCACAGGCTGAGCTGGG 375  
 Qy 626 AACCGTATCATCGCGATCCGGAGT 649  
 Db 376 AGCCCTCCATGCGAACCCTGAGT 399  
 RESULT 15  
 AAH05727  
 ID AAH05727 standard; cDNA; 547 BP.  
 XX  
 AC AAH05727;  
 XX  
 DT 26-JUN-2001 (first entry)  
 XX  
 DE Human cDNA clone (5'-primer) SEQ ID NO:2562.  
 XX  
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN EP1074617-A2.  
 XX  
 PD 07-FEB-2001.  
 XX  
 PF 28-JUL-2000; 2000EP-0116126.  
 XX  
 PR 29-JUL-1999; 99JP-0248036.  
 PR 27-AUG-1999; 98JP-0200253.  
 PR 11-JAN-2000; 2000JP-0118776.













```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/16216
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/353,700
; FILING DATE: 09-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E. 36,252
; REGISTRATION NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10136 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; PCT-US95-16216-2

Query Match 3.6%; Score 56.2; DB 5; Length 10136;
Best Local Similarity 62.4%; Pred. No. 1.7e-05;
Matches 88; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 1415 CTGGTGGCAGCAGCAAGTTAAGCTATGAGCTATGAGCTATTAACAATCATTAACCTGTGACA 1474
DB 9990 CTGTTAGCAATGCCATTCCTCTACTGCAATGTAATAGTATTAAGCTATGTATATAAGC 10049
QY 1475 CGGTAAGAAAATCTTAGGAATAAGTAAACGCTCTGAGTAAAAAAGGAAAAAAGAAAAA 1534
DB 10050 TTTTGGTAATATGTTACAATTAATAAGTACAGCAGCAGTATATAAAAAAAGAAAAA 10109
QY 1535 AAAAAAAGAAAAAAGAAAAA 1555
DB 10110 AAAAAAAGAAAAAAGAAAAA 10130

RESULT 9
US-09-801-052-1
; Sequence 1, Application US/09801052
; Patent No. 6368842
; GENERAL INFORMATION:
; APPLICANT: BEASLEY, Ellen
; TITLE OF INVENTION: ISOLATED HUMAN PHOSPHOLIPASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PHOSPHOLIPASE
; TITLE OF INVENTION: PROTEINS, AND USES THEREOF
; FILE REFERENCE: CL001045
; CURRENT APPLICATION NUMBER: US/09/801.052
; CURRENT FILING DATE: 2001-03-08
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1872
; TYPE: DNA
; ORGANISM: Human
US-09-801-052-1
Query Match 3.6%; Score 56; DB 4; Length 1872;
Best Local Similarity 64.8%; Pred. No. 8e-06;
Matches 83; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 1428 CAAGTTATTAAGCTATGGCTTATTAACAATCATTAACCTGACACGGTAAGAAAAATA 1487
DB 1720 CCAGGCAAAAGGCCGCCAGGGTTATTAATAGTAAATCACTTGTCTGTAAAAAAGAAAAA 1779

QY 1488 CTTAGGAATAAGTAAACGCTCCTGAGCTAAAAAAGAAAAAAGAAAAAAGAAAAA 1547
DB 1780 AAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 1839
QY 1548 AAAAAAAG 1555
DB 1840 AAAAAAAG 1847

RESULT 10
US-08-283-917-8
; Sequence 8, Application US/08283917
; Patent No. 5849557
; GENERAL INFORMATION:
; APPLICANT: ADACHI, HIDEKI
; APPLICANT: TSUJIMOTO, MASAFUMI
; APPLICANT: INOUE, KEIZO
; APPLICANT: ARAY, HIROYUKI
; TITLE OF INVENTION: OXIDIZED PHOPHOLIPID DEGRADING ENZYME
; TITLE OF INVENTION: AND GENE THEREOF
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER &
; ADDRESS: NEUSTADT, P.C.
; STREET: 1755 S. Jefferson Davis Highway, Suite 400
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/283,917
; FILING DATE: 03-AUG-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 209943/1993
; FILING DATE: 03-AUG-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Oblon, No. 5849557man F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 2292-030-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 413-3000
; TELEFAX: (703) 413-2220
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2085 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: CDNA
; ORIGINAL SOURCE: Bos taurus
; ORGANISM: Bos taurus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 844..2073
US-08-283-917-8
Query Match 3.6%; Score 55.8; DB 2; Length 2085;
Best Local Similarity 46.7%; Pred. No. 9.6e-06;
Matches 177; Conservative 0; Mismatches 202; Indels 0; Gaps 0;

QY 1025 AGTCGCTGTTTCCCTGTTGGATGACAGACCCCTCTATCTGTGGCGGACACAGAACCA 1084
DB 1331 AGCTCTGCTTCATGTTCTGCAGATATGACCATTAAGCTATGAGGATTTTCAGGGCTTTG 1390
QY 1085 AGTCGCTTGAGCGGCATGACAGGCGACAGCAACGTCGTAACGATGTGAAATATTTCGCCG 1144
```

1391 AATGATCAGAACCATGATGCGCATGACCAACAATGTTTTCAGTAGCCATCATGCCCA 1450  
1145 ATGTAAGCTAATTCGCTGCTTCATTTGACAAGTCAGTGGTCTGTGCGGAGCCAGG 1204  
1451 ATGGAGATCATATAGTGTCTGCTCAAGGATATAAAATGTGGAAATGCGCAAA 1510  
1205 ATGGTCAGTACATGCCACCTTCGCGGTCATGTGCGAGGCTTTTACACGGTTTGCCTGGT 1264  
1511 CTGGCTACTGTGGAAGACATTCACAGGACACAGAGAAATGGGTACGTATGGTGGCGCAA 1570  
1265 CCGGAGCTCCGCTGATGTTTTCGCGGAGCAAGAACTCAACTTAAGAATGAGTG 1324  
1571 ATCAAGACGGACTCTGATAGCCAGCTGTTCANAGACACACTGTGCGGTATGGGTGG 1630  
1325 TGCAAGAGAAAGTGGCAGAGAGTGCCTGGACATGCGGATGAGGTGTCGGAGTGG 1384  
1631 TAGCAACAAGAAATGCAAGCTGAGCTTCGAGAAACATGAGCATGTGGTAAATGCATTT 1690  
1385 ACTGGGCGCCGATGGCTC 1403  
1691 CTTGGGCTCTGAAAGCTC 1709

## RESULT 11

US-08-961-716-8  
; Sequence 8, Application US/08961716  
; Patent No. 5880272  
; GENERAL INFORMATION:  
; APPLICANT: ADACHI, HIDEKI  
; APPLICANT: TSUJIMOTO, MASAFUMI  
; APPLICANT: INOUE, KEIZO  
; APPLICANT: ARAI, HIROYUKI  
; TITLE OF INVENTION: OXIDIZED PHOPHOLIPID DEGRADING ENZYME  
; TITLE OF INVENTION: AND GENE THEREOF  
; NUMBER OF SEQUENCES: 31  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MATER &  
; ADDRESSEE: NEUSTADT, P.C.  
; STREET: 1755 S. Jefferson Davis Highway, Suite 400  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22202  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/961,716  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/283,917  
; FILING DATE: 03-AUG-1994  
; APPLICATION NUMBER: JP 209943/1993  
; FILING DATE: 03-AUG-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Oblon, No. 5880272man F.  
; REGISTRATION NUMBER: 24,618  
; REFERENCE/DOCKET NUMBER: 2292-030-0  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 413-3000  
; TELEFAX: (703) 413-2220  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2085 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: cDNA

; ORIGINAL SOURCE:  
; ORGANISM: Bos taurus  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 844..2073  
US-08-961-716-8  
Query Match 3.6%; Score 55.8; DB 2; Length 2085;  
Best Local Similarity 46.7%; Pred. No. 9.6e-06;  
Matches 177; Conservative 0; Mismatches 202; Indels 0; Gaps 0;  
QY 1025 AGTCGCTGTTTCCTGTCGATGACACACCTCTATCTGTGCGGACACACCAACA 1084  
Db 1331 AGTCTTGCTTCATGTTCTGCGATATGACCAATTAAGCTATGGATTTTCAGGCTTTG 1390  
QY 1085 AGTGGCTTGAGCGGATGACAGGACACAGAGTGGTCAAGGATGGAATATTCGCGG 1144  
Db 1391 AATGCATCAGAACCATGATGCGCATGACCAATGTTTCTTCAGTAGCCATCATGCCCA 1450  
QY 1145 ATGTAAGCTAATTCGCTGCTTCATTTGACAAGTCAGTGGCTGTGCGGAGCCAGG 1204  
Db 1451 ATGGAGATCATATAGTGTCTGCTCAAGGATATAAAATGTGGAAATGCGCAAA 1510  
QY 1205 ATGTCAGTACATGCGGACCTTCGCGGTCATGTCAGGCTGTTTACAGGTTGCTGCT 1264  
Db 1511 CTGGCTACTGTGTAAGACATTCACAGGACACAGAGATGGGTACGTATGGTGGCGCAA 1570  
QY 1265 CCGGAGCTCCGCTGATGTTTTCGCGGAGCAAGAACTCAACTCTAAAGATGAGTG 1324  
Db 1571 ATCAAGACGGCACTCTGATAGCCAGCTGTTCANAGACACACTGTGCGGTATGGGTGG 1630  
QY 1325 TGCAAGAGAAAGTGGCAGAGAGTGCCTGGACATGCGGATGAGGTGTCGGAGTGG 1384  
Db 1631 TAGCAACAAGAAATGCAAGCTGAGCTTCGAGAAACATGAGCATGTGGTAAATGCATTT 1690  
QY 1385 ACTGGGCGCCGATGGCTC 1403  
Db 1691 CTTGGGCTCTGAAAGCTC 1709

## RESULT 12

US-09-182-816-22  
; Sequence 22, Application US/09182816  
; Patent No. 6143542  
; GENERAL INFORMATION:  
; APPLICANT: Wisniewski, Nancy  
; APPLICANT: Silver, Gary M.  
; APPLICANT: Lo, Katherine C.  
; APPLICANT: Brandt, Kevin S.  
; TITLE OF INVENTION: NOVEL FLEA EPOXIDE HYDROLASE NUCLEIC ACID MOLECULES,  
; TITLE OF INVENTION: PROTEINS AND USES THEREOF  
; FILE REFERENCE: FC-3-CI  
; CURRENT APPLICATION NUMBER: US/09/182,816  
; CURRENT FILING DATE: 1998-10-29  
; EARLIER APPLICATION NUMBER: 08/989,510  
; EARLIER FILING DATE: 1997-12-12  
; NUMBER OF SEQ ID NOS: 31  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 22  
; LENGTH: 1736  
; TYPE: DNA  
; ORGANISM: Ctenocephalides felis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (159)..(1553)  
US-09-182-816-22

Query Match 3.6%; Score 55.4; DB 3; Length 1736;  
Best Local Similarity 64.3%; Pred. No. 1.1e-05;  
Matches 83; Conservative 0; Mismatches 46; Indels 0; Gaps 0;  
QY 1427 ACAAGATTATAAGCTATGGCTTATTAACAAATCAATTAATCTGTACACGGTAAGAAAT 1486

Db 1588 ATAATGTTAAATAAATCTAATTAATCTGTAATAACGATATGGAATTTTATTTCAAACT 1647  
QY 1487 ACTTAGGAATAAGTAAGTAAACGCTCTGAGTAATAAAAAAAAAAAAAAAAAAAAAA 1546  
Db 1648 TGTCAATATATAA 1707  
QY 1547 AAAAAAAAAA 1555  
Db 1708 AAAAAAAAAA 1716

RESULT 13  
US-09-182-816-24/c  
; Sequence 24, Application US/09182816  
; Patent No. 6153397  
; GENERAL INFORMATION:  
; APPLICANT: Wisniewski, Nancy  
; APPLICANT: Silver, Gary M.  
; APPLICANT: Lo, Katherine C.  
; APPLICANT: Brandt, Kevin S.  
; TITLE OF INVENTION: NOVEL FLEA EPOXIDE HYDROLASE NUCLEIC ACID MOLECULES,  
; FILE REFERENCE: FC-3-C1  
; CURRENT APPLICATION NUMBER: US/09/182,816  
; CURRENT FILING DATE: 1998-10-29  
; EARLIER APPLICATION NUMBER: 08/989,510  
; EARLIER FILING DATE: 1997-12-12  
; NUMBER OF SEQ ID NOS: 31  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 24  
; LENGTH: 1736  
; TYPE: DNA  
; ORGANISM: Ctenocephalides felis  
US-09-182-816-24

Query Match 3.6%; Score 55.4; DB 3; Length 1736;  
Best Local Similarity 64.3%; Pred. No. 1.le-05;  
Matches 83; Conservative 0; Mismatches 46; Indels 0; Gaps 0;  
QY 1427 ACAAGTTATAAGCTATGGCTTATTAAACAATCAATTAACCTGTACACGTAAGAAAT 1486  
Db 149 ATAATGTTAAATAAATAATGTAATTAATCTGTGAATAAACAATATGGAATTTTATTTCAAACT 90  
QY 1487 ACTTAGGAATAAGTAAACGCTCTGAGTAATAAAAAAAAAAAAAAAAAAAAAA 1546  
Db 89 TGTCAATATATAA 30  
QY 1547 AAAAAAAAAA 1555  
Db 29 AAAAAAAAAA 21

RESULT 14  
US-09-471-528-22  
; Sequence 22, Application US/09471528  
; Patent No. 6153397  
; GENERAL INFORMATION:  
; APPLICANT: Wisniewski, Nancy  
; APPLICANT: Silver, Gary M.  
; APPLICANT: Lo, Katherine C.  
; APPLICANT: Brandt, Kevin S.  
; TITLE OF INVENTION: FLEA EPOXIDE HYDROLASE PROTEINS AND USES THEREOF  
; FILE REFERENCE: FC-3-C1-1  
; CURRENT APPLICATION NUMBER: US/09/471,528  
; CURRENT FILING DATE: 1999-12-27  
; EARLIER APPLICATION NUMBER: 09/182,816  
; EARLIER FILING DATE: 1998-10-29  
; EARLIER FILING DATE: 1997-12-12  
; NUMBER OF SEQ ID NOS: 35  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 22  
; LENGTH: 1736

; TYPE: DNA  
; ORGANISM: Ctenocephalides felis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (159)..(1553)  
US-09-471-528-22  
Query Match 3.6%; Score 55.4; DB 3; Length 1736;  
Best Local Similarity 64.3%; Pred. No. 1.le-05;  
Matches 83; Conservative 0; Mismatches 46; Indels 0; Gaps 0;  
QY 1427 ACAAGTTATAAGCTATGGCTTATTAAACAATCAATTAACCTGTACACGTAAGAAAT 1486  
Db 1588 ATAATGTTAAATAAATAATGTAATTAATCTGTGAATAAACAATATGGAATTTTATTTCAAACT 1647  
QY 1487 ACTTAGGAATAAGTAAACGCTCTGAGTAATAAAAAAAAAAAAAAAAAAAAAA 1546  
Db 1648 TGTCAATATATAA 1707  
QY 1547 AAAAAAAAAA 1555  
Db 1708 AAAAAAAAAA 1716

RESULT 15  
US-09-471-528-24/c  
; Sequence 24, Application US/09471528  
; Patent No. 6153397  
; GENERAL INFORMATION:  
; APPLICANT: Wisniewski, Nancy  
; APPLICANT: Silver, Gary M.  
; APPLICANT: Lo, Katherine C.  
; APPLICANT: Brandt, Kevin S.  
; TITLE OF INVENTION: FLEA EPOXIDE HYDROLASE PROTEINS AND USES THEREOF  
; FILE REFERENCE: FC-3-C1-1  
; CURRENT APPLICATION NUMBER: US/09/471,528  
; CURRENT FILING DATE: 1999-12-27  
; EARLIER APPLICATION NUMBER: 09/182,816  
; EARLIER FILING DATE: 1998-10-29  
; EARLIER FILING DATE: 1997-12-12  
; NUMBER OF SEQ ID NOS: 35  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 24  
; LENGTH: 1736  
; TYPE: DNA  
; ORGANISM: Ctenocephalides felis  
US-09-471-528-24

Query Match 3.6%; Score 55.4; DB 3; Length 1736;  
Best Local Similarity 64.3%; Pred. No. 1.le-05;  
Matches 83; Conservative 0; Mismatches 46; Indels 0; Gaps 0;  
QY 1427 ACAAGTTATAAGCTATGGCTTATTAAACAATCAATTAACCTGTACACGTAAGAAAT 1486  
Db 149 ATAATGTTAAATAAATAATGTAATTAATCTGTGAATAAACAATATGGAATTTTATTTCAAACT 90  
QY 1487 ACTTAGGAATAAGTAAACGCTCTGAGTAATAAAAAAAAAAAAAAAAAAAAAA 1546  
Db 89 TGTCAATATATAA 30  
QY 1547 AAAAAAAAAA 1555  
Db 29 AAAAAAAAAA 21

Search completed: June 2, 2003, 06:54:25  
Job time : 104 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 2, 2003, 04:56:45 ; Search time 219 seconds  
(without alignments)

9574.008 Million cell updates/sec

Title: US-09-830-980-2

Perfect score: 1555

Sequence: 1 aattcccaaaaatgcaggga.....aaaaaaaaaaaaaaaaaaaa 1555

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 845702 seqs, 674182571 residues

Total number of hits satisfying chosen parameters: 1691404

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published\_Applications\_NA:

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq.\*
- 11: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*
- 12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq.\*
- 13: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*
- 14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	478	30.7	1859	9	US-10-132-744A-5
2	466.2	30.0	1928	9	US-10-132-744A-1
3	249.4	16.0	1170	9	US-10-132-744A-3
4	155.8	10.0	1722	9	US-09-774-639-34
5	155.8	10.0	1722	9	US-09-969-730-87
6	133.4	8.6	418	9	US-09-918-995-11539
7	120	7.7	445	9	US-09-918-995-37909
8	79.4	5.1	3025	10	US-09-764-853-29
9	59.6	3.8	1023	10	US-09-794-257-4
10	59.4	3.8	2431	10	US-09-925-300-464
11	59	3.8	2674	9	US-10-003-295-1
12	58.4	3.8	453	9	US-09-918-995-30837
13	57.6	3.7	1602	9	US-09-997-279-19
14	57.6	3.7	1602	10	US-09-813-358-19
15	57.4	3.7	393	10	US-09-960-352-4582
16	57.4	3.7	1221	9	US-10-237-381-2
17	57.4	3.7	1241	9	US-09-822-846-582
18	57	3.7	266	9	US-10-091-483-42
19	57	3.7	266	10	US-09-764-846-42

20	56.8	3.7	1378	9	US-10-098-841-121
21	56.8	3.7	1522	9	US-10-098-841-122
22	56.8	3.7	3466	9	US-10-042-141-35
23	56.8	3.7	3466	10	US-09-726-643-35
24	56.4	3.6	447	9	US-09-918-995-13767
25	56.4	3.6	475	9	US-09-918-995-31138
26	56.4	3.6	1133	9	US-10-282-048-1
27	56.4	3.6	3449	9	US-09-925-299-225
28	56.4	3.6	3449	10	US-09-925-299-225
29	56.2	3.6	347	10	US-09-960-352-1096
30	56.2	3.6	2732	9	US-09-759-056-1
31	56.2	3.6	2732	9	US-09-901-812-1
32	56.2	3.6	2732	9	US-10-237-884-79
33	56.2	3.6	2732	9	US-10-230-163-79
34	56.2	3.6	2732	9	US-10-218-631-79
35	56.2	3.6	2732	9	US-10-230-338-79
36	56.2	3.6	2732	9	US-10-230-414-79
37	56.2	3.6	2732	9	US-10-218-849-79
38	56.2	3.6	2732	9	US-10-218-159A-79
39	56.2	3.6	2732	9	US-10-227-873-79
40	56.2	3.6	2732	9	US-10-227-883-79
41	56.2	3.6	2732	9	US-10-219-076-79
42	56.2	3.6	2732	9	US-10-230-434-79
43	56.2	3.6	2732	9	US-10-219-003-79
44	56.2	3.6	2732	9	US-10-219-075-79
45	56.2	3.6	2732	9	US-10-219-464-79

#### ALIGNMENTS

##### RESULT 1

US-10-132-744A-5  
; Sequence 5, Application US/10132744A  
; Publication No. US20030027261A1

; GENERAL INFORMATION:  
; APPLICANT: Utku, Nalan

; TITLE OF INVENTION: No. US20030027261A1el genes rZap7/A, rZap7/B and rZap7 involve activation and uses thereof  
; FILE REFERENCE: Utku-4 CON

; CURRENT APPLICATION NUMBER: US/10/132.744A  
; CURRENT FILING DATE: 2002-04-24

; PRIOR APPLICATION NUMBER: PCT/EP00/10670  
; PRIOR FILING DATE: 2000-10-30

; PRIOR APPLICATION NUMBER: 60/185,016  
; PRIOR FILING DATE: 2000-02-25

; PRIOR APPLICATION NUMBER: 60/162,675  
; PRIOR FILING DATE: 1999-11-01

; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: Patentin Ver. 2.1

; SEQ ID NO 5

; LENGTH: 1859

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: miscellaneous feature

; LOCATION: 1154; 379

; OTHER INFORMATION: ATCG; variable amino acid

US-10-132-744A-5

Query Match

Best Local Similarity 59.8%; Score 478; DB 9; Length 1859;

Matches 837; Conservative 0; Mismatches 556; Indels 6; Gaps 2;

QY 54 GATACAGCGCGCTCGTTTACACGGCGAGAGCGCGCGCGCAATCGACCTGCGCGG 113

DB 70 GCTAGTGCAGTTTCAGGATGAGCGGCGGCGAGCTGCTGGGTTCCCGTTCGACGTGCGCGT 129

QY 114 AGGAATCACTACCCAGCAATGGGACTGATTTCACACGCGCTGCTGAAACAGGAGGAGC 173

DB 130 GGACATCACCCCGGACAGCTGCAGCTCGTGTCAACGCGCTACTTGGCCCGGAGGATCC 189

QY 174 CACTCCATATTGTTTTCGTGGCGAGGATGAGATCAAGAGACCTTGAGGACACGTT 233



```
Db 499 AGTCCGCTTTGGGATCTCAGCACAGAAACTCCACATTTTACATCTAAAGGGCATACACA 558
QY 474 GTGGGTTCTGCGGTATCTCGGCTCGGATGCAAAAGTTGGCCAGCGGTTCGAAAGC 533
Db 559 CTGGGTTCTCAGTATGTTGGTCTCCAGATGCAAAAGTTGGCTCAGGATGTAAGA 618
QY 534 GGGGCTATAATCTCTGGAGACCGGATGCAAGAGGGGCGGACCCCTTGAGTGG 593
Db 619 TAGTCAGATCTTCAATTTGGGACCAAGCACAGGGAAGACAGATTTGGCAAACTAAACAGG 678
QY 594 GCACAGAAACATCAACTGCTCGCTGGGACCGATATCATCGCATCCGGAGTCCAG 653
Db 679 GCATCAAGTGAATACATGCTGTGTGGGAACCTCTCCACCTGAACCCAGAGAGCG 738
QY 654 GAAACTTGCTCGGCGATGGAGAGCGGGACTCGCGATTTGGGACGTAAATTTGGCCCA 713
Db 739 ATACCTAGCCAGTGCTCCAGCGCGCGTGGACCGGATCTGGACAACTTGCAGGCGG 798
QY 714 GTGCCCTTATGAACATTTGCCGACACACAAATGCTGTGACAGCAGTGAGATGGGTGGAGC 773
Db 799 CTGTGAGCGCATCTCACCGGGCACACCCAGTCGGTCACTGTCTCCGTTGGGAGGGGA 858
QY 774 GGGGCTTATTTATACATCTCCAAAGATCGACAGTGAAGATGTGGCAGCAGTGTATGG 833
Db 859 CGGGCTTCTTACTCTGCTCCAGGACCGCACCTCAAGTCTGGAAGTCTGGAGCTCATGCGG 918
QY 834 AATCTTGTCCGGACGTTCTCTGGCCAAAGCTCACTGGGTAAACACATTTGGCTGAGCAC 893
Db 919 TGTGCTGTCCGGACTCTGCAAGGCCACGGCCACTGGGTGAACACCATGGCCCTCAGCAC 978
QY 894 CGATTAGCTCTCGCACGTGTCATTCATTCGCGTGAAGGATCGTCCCAAGAGCC---A 950
Db 979 TGATATGCCCCTCGCACGTGGGGCTTTGAACCTGCTGAGGCCCTCAGTTAATCCCAAGA 1038
QY 951 CCTCAGTTGAGCACTGAGGAATTCGAGGAATTCGCTTGAAGCGGTACCGCGCGGTG 1010
Db 1039 CTTCAAGGATCTCTGACAGGATTTGAAGGAGAGGGCTCTGAGCCGATACAACTGCTGCG 1098
QY 1011 CCCTGAGAGGTGAGTCTGCTGTTTCTGTTCCGTGATGACACACCCCTTATCTGTGG-- 1068
Db 1099 GGGCCAGGTCAGAGAGGCTGTGTCTGGCTCGGACGACTTCACCTTATCTCTGTGTC 1158
QY 1069 -CGGNAACCAACAGCAAGTGGTTGAGCGCATGACAGGCGCACCAAGCTGGTCAACGA 1127
Db 1159 CCCAGCAGAGGACAAAGCCCTCACTCGGATGACAGGACCAAGCTCTCATCAACA 1218
QY 1128 TGTGAAATATTCCCGGATGTAAAGCTAATTGCGTCTGCTTCATTTGACAACTCAGTGG 1187
Db 1219 GGTGCTCTCTCTCTGACTCCCGCATCGTGGCTAGTGGCTCTCTTGACAAGTCCATCAA 1278
QY 1188 TGTGTGGCGAGCGATGGTCACTAGTACATGCCACCTTCGGGGTCAATGTCAGGCTGT 1247
Db 1279 GCTGTGGGATGGAGAGGGGCAAGTACCTGGCTTCCCTACGCGGCCAGCTGGCTGCCGT 1338
QY 1248 TTACACGTTGCTGTGGTCCGGGACTCCCGCTGTGATTTGTTTCGGCAGCAAGACTCAAC 1307
Db 1339 GTACCAGATTTCGTTGGTCACTGACAGTGGCTCGCTGGTTCAGCGGCGAGCAGTGAAGCAG 1398
QY 1308 TCTAAAGTATGGAGTGTGCAGACGAAGAACTGGCAGCAGGAGTGTGCTGGACATGCGGA 1367
Db 1399 ACTGAAGTGTGGGATGTGAAGGCCCAAGAGCTGGCCATGAGACCTGCGCGGCCAGCGGA 1458
QY 1368 TGAGTGTTCGGAGTGGAGTGGCGGCCCGATGGCTCTAGAGTTGCTCTGGTGGCAAGA 1427
Db 1459 TGAGGTATATGCTGTGACTGGAGTCCAGATGGCCAGAGAGTGGCAAGTGGTGGGAAGA 1518
QY 1428 CAAAGTTATAAGCTATGG 1446
Db 1519 CAANTGCCTCCGGATATGG 1537
```

RESULT 3

US-10-132-744A-3

```
; Sequence 3, Application US/10132744A
; Publication No. US20030027261A1
; GENERAL INFORMATION:
; APPLICANT: Utku, Nalan
; TITLE OF INVENTION: No. US20030027261A1el genes Tzap7/A, Tzap7/B and Tzap7 involve
; TITLE OF INVENTION: activation and uses thereof
; FILE REFERENCE: Utku-4 CON
; CURRENT APPLICATION NUMBER: US/10/132,744A
; PRIOR FILING DATE: 2002-04-24
; PRIOR APPLICATION NUMBER: PCT/EP00/10670
; PRIOR FILING DATE: 2000-10-30
; PRIOR APPLICATION NUMBER: 60/185,016
; PRIOR FILING DATE: 2000-02-25
; PRIOR APPLICATION NUMBER: 60/162,675
; PRIOR FILING DATE: 1999-11-01
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 1170
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3)..(785)
US-10-132-744A-3
```

```
Query Match 16.0%; Score 249.4; DB 9; Length 1170;
Best Local Similarity 59.3%; Pred. No. 7.6e-57;
Matches 462; Conservative 0; Mismatches 311; Indels 6; Gaps 2;
```

```
QY 674 GAGACGGGACTTGGCGGATTTGGGACGTAAATTTGGCCAGTGCCTTATGACATTTGCCG 733
Db 1 GCGGCGCGCTGCACCGGATCTGGGACACAACTGCAGCGCGCTGTGAGCGCATCCTCACCG 60
QY 734 GACACAAATCTCTGTGACAGTGTGAGTGGGTGGAGCGGCGCTTATTTATACATCT 793
Db 61 GGACACACCACTGCTGCTACCTGCTCCGGTGGGAGGGGACGGGCTTCTTACTCTGCT 120
QY 794 CCAAGATGCGACATGAAGATGTGGCGGACGACTGATGGAATCTGTGCGGACGTTCT 853
Db 121 CCAGAGACCGCACCATCAAGTCTGGAGAGCTCATACGGTGTGCTGTGCGGAGCTTGC 180
QY 854 CTGGCCAACTCACTGGGTAAACAACTTCGCTGAGCACCAGTATCTCTGCGGCACTG 913
Db 181 AAGGCCACGGCACTGGGTGAACACCATGGCCCTAGCCTGACTGACTATGCCCTGGCACTG 240
QY 914 GTCCATTCATCGGTGAGGATCGCTCCAGAGCC---ACCTCAGTTTGTAGCAGTCTGAGG 970
Db 241 GGGCCCTTGAACCTGCTGAGGCTCAGTTAATCCCAAGACCTCCAAGGATCTTTCAGG 300
QY 971 AATTGAGGAATCTGCTTGAAGCGCTACCGCGCGTGTGCGCTGACGAGTGGAGTCCG 1030
Db 301 AGTTGAAGGAGAGGGCTCTGAGCGGATACAACTCTGCGGGGCCAGGGTCCAGAGAGGC 360
QY 1031 TGGTTTCTTGTGGATGACAAACCTCTATCTGTGG---CGGAACACACAGAACAACT 1087
Db 361 TGGTGTCTGCTCCGACACTTCACCTTATTCCTGTGGTCCCGGCGGCGGAGGAGG 420
QY 1088 CGGTGAGCGCATGACAGGCGCACAGAACGTGGTCAACGATGTGGAATATTCGCGGATG 1147
Db 421 CTCTCACTCGGATGACAGGACCAACAGCTCTCATCAACAGGTGCTTCTCTCTCTGACT 480
QY 1148 TAAAGCTTAATTCGCTGCTTCATTTGACAGTCAAGTCAAGTCTGTGCGGAGCAGCAGATG 1207
Db 481 CCGCATCTGTGGCTAGTGGCTCTCTTTTGACAGTCCATCAAGCTGTGGGATGGCAGACGG 540
QY 1208 GTACATACATGGCCACCTTCCGGGGTCAATGTGAGGCTGTGTACAGGTTGTGCGGTGCTCG 1267
Db 541 GCAAGTACCTGGCTTCCCTACGCGGCCACGTGGCTGCGGTGTACCAAGATTTGGTGGTCA 600
QY 1268 CGGACTCCCGCTTGTATTGTTCCGGCAGCAAAAGACTCAACTCTTAAAGTATGGAGTGTGC 1327
Db 601 CTGACATCGGCTCCTGGTTCAGCGCGCAGCAGTGCAGCAGCAGTGAAGGTGTGGATGTGA 660
```

```

1328 AGACGAAGAACTGGCAGAGAGCTGCTGGACATCGGATGAGTGTGCGAGTGGACT 1387
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
661 AGGCCAGAGAGCTGGCCATGACCTGCCCGCCACCGCGATGAGTATATGCTGTGACT 720
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1388 GGGCGCCGATGGCTCTAGAGTGTGCTTCTGGTGGCAAGGACAAAGTTATTAAGCTATGG 1446
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
721 GGATCCAGATGGCCAGAGAGTGGCAAGTGGTGGGAAGGACAAATGGCTCCGGATATGG 779
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 4
US-09-774-639-34
; Sequence 34, Application US/09774639
; Publication No. US20030003555A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 90 Human Secreted Proteins
; FILE REFERENCE: P2013P1
; CURRENT APPLICATION NUMBER: US/09/774,639
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/244,112
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-02-04
; NUMBER OF SEQ ID NOS: 371
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 34
; LENGTH: 1722
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (2)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (413)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-774-639-34

Query Match 10.0%; Score 155.8; DB 9; Length 1722;
Best Local Similarity 58.4%; Pred. No. 1.4e-31;
Matches 326; Conservative 1; Mismatches 224; Indels 7; Gaps 3;

QY 896 APTAGCTTCGCGACTGGTCCATTCATCCGCTGAGGAGAGAGGCTCTGAGCGGATACAACTCGTGGCGG 952
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3 ACTATGCCCTGGCGACTGGGCGCTTTGAACCTCTGAGGCGCTTCAGTTAAATCCCAAGACC 62
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 953 TCAGTTTGGACACTGAGGAATCTGCCTTGAAGCGCTTACCAGCGCGTGTGCC 1012
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 63 TCCAAGGATCCTTGCAAGAGTTGAAGGAGAGGCTCTGAGCGGATACAACTCGTGGCGG 122
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1013 CTGACGAGGTGAGTGGCTGTTTCTTTCGATGATCAACACCTCTATCTGTGG--C 1069
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 123 GCCAGGTCCAGAGGCTGGTGTCTGGTCCGACGCTTACCTTATCTGTGTGTC 182
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1070 GGAACACACCAAGAGTGGTGTGAGCGCATGACAGGCGACAGCAAGCTGGTCAACGATG 1129
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 183 CAGCAGAGGRCAAAAGCCTCTCAGTCGGATGACAGGACACAAAGCTCTCATCAACGAGG 242
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1130 TGAATATTCCCGGATGTAAGCTAATTCGCTCTCTTTCATTTGACAAGTCAAGTCGCTC 1189
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 243 TGCTCTTCTCTGACTCCCGCATCGTGGCTAGTGCCTCTTTGACAAGTCCATCAAGC 302
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1190 TGTGGCAGCGAGCGATGGTCAATACATGCGCACCTTCGGGGTCATGTCAGGCTGTTT 1249
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 303 TGTGGATGCGAGCGGCGCATGCTTCCCTACGCGGCGACGCTGGTGGCGGTG 362
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1250 ACACGGTTCGCTGGTCCGCGGACTCCGCTTGTATGTTTCCGGCAGCA--AAGACTCAACT 1308
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 363 ACCAGATTGGTGTGCTGACATGCTGGTCTGCTGGTTCAGCGCGCAGCATGACAGCACA 422
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1309 CTAAAGTATGAGTGTGACAGCAAGAAATGCGACAGAGCTGCTGGACATGCGGAT 1368
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 423 CTGAGGTGTGGATGTGAAGGCCCAAGAGCTGGCCATGGACCTGCCGCGCCAGCGGAT 482
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1369 GAGTGTTCGGATGGACTGGCGCGCGATGGCTCTAGAGTTGCTTGGTGGCAAGGAC 1428
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 483 GAGTATATGCTGTTACTGGAGTCCAGATGGCCAGAGTGGCAAGTGGTGGGAAGGAC 542
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1429 AAAGTTATAAGCTATGG 1446
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 543 AATGCTTCGGATATGG 560
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 5
US-09-969-730-87
; Sequence 87, Application US/09969730
; Publication No. US20030054443A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 90 Human Secreted Proteins
; FILE REFERENCE: P2013P2
; CURRENT APPLICATION NUMBER: US/09/969,730
; CURRENT FILING DATE: 2001-10-04
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: 09/774,639
; PRIOR FILING DATE: 2001-02-01
; PRIOR APPLICATION NUMBER: 60/238,291
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 09/244,112
; PRIOR FILING DATE: 1999-02-04
; PRIOR APPLICATION NUMBER: PCT/US98/16235
; PRIOR FILING DATE: 1998-08-04
; PRIOR APPLICATION NUMBER: 60/056,371
; PRIOR FILING DATE: 1997-08-19
; PRIOR APPLICATION NUMBER: 60/056,732
; PRIOR FILING DATE: 1997-08-19
; PRIOR APPLICATION NUMBER: 60/056,366
; PRIOR FILING DATE: 1997-08-19
; PRIOR APPLICATION NUMBER: 60/056,364
; PRIOR FILING DATE: 1997-08-19
; PRIOR APPLICATION NUMBER: 60/056,370
; PRIOR FILING DATE: 1997-08-19
; PRIOR APPLICATION NUMBER: 60/056,367
; PRIOR FILING DATE: 1997-08-19
; PRIOR APPLICATION NUMBER: 60/056,365
; PRIOR FILING DATE: 1997-08-19
; PRIOR APPLICATION NUMBER: 60/056,731
; PRIOR FILING DATE: 1997-08-19
; PRIOR APPLICATION NUMBER: 60/056,557
; PRIOR FILING DATE: 1997-08-19
; PRIOR APPLICATION NUMBER: 60/056,563
; PRIOR FILING DATE: 1997-08-19
; PRIOR APPLICATION NUMBER: 60/055,970
; PRIOR FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: 60/055,986
; PRIOR FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: 60/055,311
; PRIOR FILING DATE: 1997-08-05
; PRIOR APPLICATION NUMBER: 60/054,808
; PRIOR FILING DATE: 1997-08-05
; PRIOR APPLICATION NUMBER: 60/054,803
; PRIOR FILING DATE: 1997-08-05
; PRIOR APPLICATION NUMBER: 60/054,804
; PRIOR FILING DATE: 1997-08-05
; PRIOR APPLICATION NUMBER: 60/054,809
; PRIOR FILING DATE: 1997-08-05
; PRIOR APPLICATION NUMBER: 60/054,806
; PRIOR FILING DATE: 1997-08-05
; PRIOR APPLICATION NUMBER: 60/055,310
; PRIOR FILING DATE: 1997-08-05
; PRIOR APPLICATION NUMBER: 60/054,798
; PRIOR FILING DATE: 1997-08-05
; PRIOR APPLICATION NUMBER: 60/055,309
; PRIOR FILING DATE: 1997-08-05
; PRIOR APPLICATION NUMBER: 60/055,312
; PRIOR FILING DATE: 1997-08-05
; PRIOR APPLICATION NUMBER: 60/054,807
; PRIOR FILING DATE: 1997-08-05
```





```
Db      63  GAGCTCATGACGGTGTGCTGTGCGGACTCTGCAAGGCCACGGCCACTGGGTGAACACCA 122
QY      881  TTGGCTGAGCAGCCGATTACGTCTGCGCACTGGTCCATTCCATCCGTTGAAGATCGCT 940
Db      123  TGGCCCTCAGCACTGACTATGCCCTTGGCACTGGGGCCCTTTGAACCTGCTGAGGCCCTAG 182
QY      941  CCAAGAGCC---ACCTCAGTTTGAAGCACTGAGGAATTGAGGAATCTGCCCTTGAAGCGCT 997
Db      183  TTATCCCAAGACCTCCAAAGGATCCTTGCGAGGATTGAAGGAGAGGCTCTGAGCCGAT 242
QY      998  ACCAGGCCGTGTGCCCTGACGAGGTGGAGTCCGTGTTCTTCTGTTGCGATGACAAACCC 1057
Db      243  ACACCTCTGTGGGGCCAGGTCACAGAGGCTGTGTCTGGCTCCGACCACTCACCT 302
QY      1058  TCTATCTGTGG---CGGAACACAGAACAGTCCGTTGAGCGCATGACAGGCAACAGCA 1114
Db      303  TATTCTGTGGTCCCGCAGCAGAGGACAAAAGGCTCTCACTCGGATGACAGCACCAAG 362
QY      1115  ACGTGTGCAACGATGTGAATATTGCGCGGATGTAAAGCTAATTGCGTCTGCTTCATTG 1174
Db      363  CTCTCATCAACAGGTGCTCTCTCTGCTGCTCCGCGATCGTGGCTAGTCCCTCTTG 422
QY      1175  ACAAGTCAGTGGCTGTGG 1194
Db      423  ACAAGTCCATCAAGCTGTGG 442
```

## RESULT 8

```
US-09-764-853-29
; Sequence 29, Application US/09764853
; Patent No. US20020090672A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PJ206
; CURRENT APPLICATION NUMBER: US/09/764,853
; PRIOR FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 939
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 29
; LENGTH: 3025
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (3011)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (3014)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (3015)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-853-29
```

```
Query Match      5.1%; Score 79.4; DB 10; Length 3025;
Best Local Similarity 49.2%; Pred.No. 6.5e-11;
Matches 209; Conservative 0; Mismatches 216; Indels 0; Gaps 0;
```

```
QY      1025  AGTCGCTGTTCTCTGTTTCGGATGACACACCCCTCTATCTGTGCGCGAACAACAGACA 1084
Db      222  AGCAACTTGCTACTGCTCTCTTGGGATACCTTTCTCATGCTATGGAATTTCAAGCCACATG 281
QY      1085  AGTGGCTTGAGCGCATGACAGGCGCAGACGTGTCACAGATGCAAAATATTCGCGG 1144
Db      282  CTAGAGCTACAGATATGTGGTCACAAGGATGTGAACAGCGTCACATTTCTCCAC 341
QY      1145  ATGTAAGCTAATTGCGTCTGCTTCATTTGACAAGTCAGTCGCTGTGCGGACGACG 1204
Db      342  ATGGAAGCTAATTGCGTCTGCTTCACGAGACAGAACCGTGAGACTCTGGAATTCCTGATA 401
```

```
QY      1205  ATGGTCAGTCATACGCGCACCTTCCGGGTCTATGTCAGGCTGTTTACACGGTTCCTGCT 1264
Db      402  AGAGAGGAAAATTTCTCAGAATTTAAAGCTCATACAGCTCCAGTTTCGAAGTGTAGACTTTT 461
QY      1265  CCGCGACTCCCGCTTGATTTTCCGGCAGCAAGACTCAACTCTCTAAAAGTATGAGGTG 1324
Db      462  CAGCTGATGGCCAGTTCTTCTAGCTACAGCTTCTGAAGACAAATCCATAAAAGTATGAGCA 521
QY      1325  TCCACAGCAAGAAACTGGCACAGGAGCTGCCTGGACATCGCGATGAGGTGTCGGAGTGG 1384
Db      522  TGTATCGCCAGCGTTCCTGTATTCCTTGTATCGACATACACATGGGTACGCTGTGCCA 581
QY      1385  ACTGGGCGCCCGATGCGTCTAGAGTTGCTCTGTGGCAAGGACAAAGTTATAAAGCTAT 1444
Db      582  AATTTTCAACCGATGGAAGACTAATTTGTCTCATGTAGTGAGGATAAACTATTATAAATTT 641
QY      1445  GGGCT 1449
Db      642  GGGAT 646
```

## RESULT 9

```
US-09-794-257-4
; Sequence 4, Application US/09794257
; Patent No. US20020009804A1
; GENERAL INFORMATION:
; APPLICANT: Meyers, Rachel
; TITLE OF INVENTION: 32705, 23224, 27423, 32700, 32712, No. US20020009804A1e1
; FILE REFERENCE: Human G-Proteins
; CURRENT APPLICATION NUMBER: US/09/794,257
; PRIOR FILING DATE: 2001-02-27
; PRIOR APPLICATION NUMBER: 60/185,606
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1023
; TYPE: DNA
; ORGANISM: homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (245)...(886)
; NAME/KEY: misc.feature
; LOCATION: (1)...(1023)
; OTHER INFORMATION: n = A,T,C or G
US-09-794-257-4
```

```
Query Match      3.8%; Score 59.6; DB 10; Length 1023;
Best Local Similarity 72.0%; Pred.No. 7.5e-06;
Matches 77; Conservative 0; Mismatches 30; Indels 0; Gaps 0;
```

```
QY      1449  TTATTAACAATCATTAACTTGTCACACGTAAGAAATATCTTAGGAATAAAGTAAACGT 1508
Db      896  TTATTTCCAAACATGCTCTCTCTACTTGAAGTGAAGAGAAATATATAGATCTT 955
QY      1509  CCTGAGTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1555
Db      956  TGTGTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1002
```

## RESULT 10

```
US-09-925-300-464
; Sequence 464, Application US/09925300
; Patent No. US20020151681A1
; GENERAL INFORMATION:
; APPLICANT: Craig Rosen,
; APPLICANT: Steve Ruben
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA101
; CURRENT APPLICATION NUMBER: US/09/925,300
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05988
```

FILE REFERENCE: 2000-03-08  
CURRENT FILING DATE: 2000-03-08  
PRIOR FILING DATE: 1999-03-12  
NUMBER OF SEQ ID NOS: 1890  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 464  
LENGTH: 2431  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-925-300-464

Query Match 3.8%; Score 59.4; DB 10; Length 2431;  
Best Local Similarity 77.4%; Pred. No. 1.4e-05;  
Matches 72; Conservative 0; Mismatches 21; Indels 0; Gaps 0;  
QY 1463 TTAACCTGTACAGGTAAGAAATACCTAGTAATAAGTAAACCTCTCTAGTAAAAA 1522  
DB 2337 TTGTTTGTATTCTGTAGCTAGTCTGCTAATAAGTAAAGATCTGAAAAA 2396  
QY 1523 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1555  
DB 2397 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2429

RESULT 11  
US-10-003-295-1  
Sequence 1, Application US/10003295  
Patent No. US20020168741A1  
GENERAL INFORMATION:  
APPLICANT: GAN, Weinlu et al.  
TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC  
ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES  
TITLE OF INVENTION: THEREOF  
FILE REFERENCE: CL001183DIV  
CURRENT FILING DATE: 2001-12-06  
NUMBER OF SEQ ID NOS: 4  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 1  
LENGTH: 2674  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-003-295-1

Query Match 3.8%; Score 59; DB 9; Length 2674;  
Best Local Similarity 55.0%; Pred. No. 1.9e-05;  
Matches 116; Conservative 0; Mismatches 95; Indels 0; Gaps 0;  
QY 1345 CAGGAGTCTGTCACATGCGATGCGGTGTCGGAGTGGAGTGGCGCCGCGATGGCTCT 1404  
DB 2439 CAGCATCCACACTGCCGCGCAGGATGCGCGCGTCTCTCTGTGTCTGCTGCTGCC 2498  
QY 1405 AGAGTGGCTCTGTGGCAGGACAAAGTATATAAGTATGGCTTATTAACAATCATT 1464  
DB 2499 AGGGTCTCTCTGCGGCGAGAAACAAATTAACACCTGTGCGCCACTGAAAAA 2558  
QY 1465 AACTGTACAGGTAAGAAATACCTAGTAATAAGTAAACCTCTCTGAGTAAAAA 1524  
DB 2559 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2618  
QY 1525 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1555  
DB 2619 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2649

RESULT 12  
US-09-918-995-30837  
Sequence 30837, Application US/09918995  
Publication No. US20030073623A1  
GENERAL INFORMATION:  
APPLICANT: Hyseq, Inc.  
TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
FROM VARIOUS CDNA LIBRARIES

FILE REFERENCE: 20411-756  
CURRENT FILING DATE: 2001-07-30  
PRIOR FILING DATE: 2001-07-30  
NUMBER OF SEQ ID NOS: 38054  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 30837  
LENGTH: 453  
TYPE: DNA  
ORGANISM: Homo sapiens  
NAME/KEY: misc\_feature  
LOCATION: (1)...(453)  
OTHER INFORMATION: n = A,T,C or G  
US-09-918-995-30837

Query Match 3.8%; Score 58.4; DB 9; Length 453;  
Best Local Similarity 48.8%; Pred. No. 1e-05;  
Matches 158; Conservative 0; Mismatches 166; Indels 0; Gaps 0;  
QY 1019 AGGTGAGTCTGCTGTTTCTGTTTCGGATGACACACCTCTATCTGTGGCGGAACCC 1078  
DB 60 ACGGGAGTGTGTAGCTCTGTTCCAGGACACAAACATCAAGCTCTGGGACATCAGGA 119  
QY 1079 AGAACAAGTGGTGTGAGCGATGACAGGACACAGAGTGTGTCAACGATGTGAATAATT 1138  
DB 120 GGAAGGCTGTGTCTTCGATACAGGGGACACAGCCGCTGCGGTGTCTCCGGTTCA 179  
QY 1139 CGCGGATGTAAGCTAATTCGCTGCTTCAATTTGACAAGTCACTGCTGTGGCGAG 1198  
DB 180 GCGCCGATGGAAGTGTGCGTGGCGCGCAGATGACACACCTGGAAGCTCTGGGATC 239  
QY 1199 CAAGCGATGTCAGTACATGCGCCACCTTCCGGGGTATGTGCGAGCTGTGTACACGGTTG 1258  
DB 240 TCACTGCCGCGAAGATGATGTCTGAGTTCCTGCTCACACGGGCTGTCAACGTGCTG 299  
QY 1259 CTGTGTCGCGGACTCCGCTTGATTTTCGGGACGAAAGACTCAACTCTAAAGTAT 1318  
DB 300 AGTTTCAACCCCAACGAGTACCTCTCGCTCCGCGACCTCTGACAGGACATCCGCTTCT 359  
QY 1319 GGAGTGTGACAGCAAGAAACTGG 1342  
DB 360 GGGACCTGGAGAAATTCAGGTGG 383

RESULT 13  
US-09-997-279-19  
Sequence 19, Application US/09997279  
Publication No. US20030059781A1  
GENERAL INFORMATION:  
APPLICANT: Chenault, Ruth A.  
APPLICANT: Xu, Jiangchun  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
OF INVENTION: AND DIAGNOSIS OF OVARIAN AND ENDOMETRIAL CANCER  
TITLE OF INVENTION: 210121.501C1  
FILE REFERENCE: 210121.501C1  
CURRENT FILING DATE: 2001-11-28  
NUMBER OF SEQ ID NOS: 230  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 19  
LENGTH: 1602  
TYPE: DNA  
ORGANISM: Homo sapien  
NAME/KEY: misc\_feature  
LOCATION: (1)...(1602)  
OTHER INFORMATION: n = A,T,C or G  
US-09-997-279-19

Query Match 3.7%; Score 57.6; DB 9; Length 1602;  
Best Local Similarity 65.1%; Pred. No. 3.3e-05;  
Matches 84; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

```
QY 1427 ACAAGTTATAAGCTATGGCTTTATTACAAATCATTAACTTGTACACGGTAAGAAAT 1486
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 188 AATATATATACATATATTGCTTTATTGGAACAATTAATATATGCTGCAATTTGAANA 247
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1487 ACTTAGGATAAGTAAGTAAACGCTCTGAGTAAATAAAAAAAAAAAAAAAAAAAAA 1546
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 248 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 307
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1547 AAAAAAAAA 1555
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 308 AAAAAAAAA 316
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
RESULT 14
US-09-813-358-19
; Sequence 19, Application US/09813358
; Patent No. US20020048759A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Pyle, Ruth
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.501
; CURRENT APPLICATION NUMBER: US/09/813,358
; NUMBER OF SEQ ID NOS: 222
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 1602
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1602)
; OTHER INFORMATION: n = A,T,C or G
US-09-813-358-19
```

```
Query Match 3.7%; Score 57.6; DB 10; Length 1602;
Best Local Similarity 65.1%; Pred. No. 3.3e-05;
Matches 84; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 1427 ACAAGTTATAAGCTATGGCTTTATTACAAATCATTAACTTGTACACGGTAAGAAAT 1486
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 188 AATATATATACATATATTGCTTTATTGGAACAATTAATATATGCTGCAATTTGAANA 247
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1487 ACTTAGGATAAGTAAGTAAACGCTCTGAGTAAATAAAAAAAAAAAAAAAAAAAAA 1546
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 248 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 307
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1547 AAAAAAAAA 1555
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 308 AAAAAAAAA 316
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
RESULT 15
US-09-960-352-4582/c
; Sequence 4582, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathalagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 4582
; LENGTH: 393
```

```
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 20-LIB3057-013-Q1-K1-E7
US-09-960-352-4582

Query Match 3.7%; Score 57.4; DB 10; Length 393;
Best Local Similarity 71.0%; Pred. No. 1.7e-05;
Matches 76; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 1449 TTATTACAAATCATTAACTTGTACACGGTAAGAAATAGTAAGTAAACGT 1508
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 233 TTATAAATAATTTTTTAAATAATAAAATTTATAAATAATAAAAAAAAAAAAA 174
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1509 CCTGAGTAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1555
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 173 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 127
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Search completed: June 2, 2003, 06:58:17
Job time : 221 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 2, 2003, 04:48:24 ; Search time 2084 Seconds  
(without alignments)  
12084.444 Million cell updates/sec

Title: US-09-830-980-2

Perfect score: 1555

Sequence: 1 aattcccaaaatgcagga.....aaaaaaaaaaaaaaaaaaaa 1555

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin.\*  
4: em\_estmu.\*  
5: em\_estov.\*  
6: em\_estpl.\*  
7: em\_estro.\*  
8: em\_hic.\*  
9: gb\_est1.\*  
10: gb\_est2.\*  
11: gb\_hic.\*  
12: gb\_est3.\*  
13: gb\_est4.\*  
14: gb\_est5.\*  
15: em\_estfun.\*  
16: em\_estom.\*  
17: gb\_gss.\*  
18: em\_gss\_hum.\*  
19: em\_gss\_inv.\*  
20: em\_gss\_pln.\*  
21: em\_gss\_vrt.\*  
22: em\_gss\_fun.\*  
23: em\_gss\_nam.\*  
24: em\_gss\_mus.\*  
25: em\_gss\_other.\*  
26: em\_gss\_pro.\*  
27: em\_gss\_rod.\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	735.6	47.3	790	9 AI518473	AI518473 ID38001.5
2	597.8	38.4	637	13 BI637470	BI637470 SD19506.5
3	591.8	38.1	725	12 BF505998	BF505998 AT08344.5
4	578.2	37.2	622	13 BI632742	BI632742 SD26605.5
5	549.2	35.3	606	13 BI232464	BI232464 R228335.5
6	548.8	35.3	626	13 BI639652	BI639652 SD22291.5

7	488.2	31.4	527	9	AI520108	AI520108 LD40082.5
8	485.2	31.2	552	13	BI635777	BI635777 SD17257.5
9	476.6	30.6	1784	11	BC035421	BC035421 Mus muscu
10	455.8	29.3	510	13	BI635228	BI635228 SD16515.5
11	424	27.3	448	9	AA735870	AA735870 GM10004.5
12	346	22.3	688	13	BM579586	BM579586 170006872
13	331.6	21.3	975	9	AL520339	AL520339 AL520339
14	314	20.2	925	9	AL526575	AL526575 AL526575
15	301.4	19.4	1114	9	AL530778	AL530778 AL530778
16	296.4	19.1	861	9	AL558947	AL558947 AL558947
17	290.4	18.7	851	9	AL558453	AL558453 AL558453
18	289.4	18.6	861	9	AL519702	AL519702 AL519702
19	284.2	18.3	927	14	BQ070427	BQ070427 AGENCOURT
20	282	18.1	997	12	BG470097	BG470097 602533586
21	278.8	17.9	939	13	BM579239	BM579239 170006872
22	278.8	17.9	939	14	BQ956834	BQ956834 AGENCOURT
23	276.6	17.8	826	9	AL526922	AL526922 AL526922
24	275.2	17.7	822	9	AL550330	AL550330 AL550330
25	268.8	17.3	1103	13	BM478420	BM478420 AGENCOURT
26	265.2	17.1	1006	13	BM474111	BM474111 AGENCOURT
27	263.6	17.0	816	9	AL527738	AL527738 AL527738
28	263.6	16.9	573	13	BM654928	BM654928 170006873
29	262.4	16.9	324	13	BI232484	BI232484 RE28559.5
30	256.6	16.5	921	13	BI833055	BI833055 603090947
31	251.4	16.2	912	9	AL522417	AL522417 AL522417
32	250.2	16.1	903	12	BG481047	BG481047 602529007
33	245.4	15.8	673	13	BM654934	BM654934 170006873
34	242.6	15.6	678	14	BQ257262	BQ257262 NISC_Ko12
35	242.6	15.6	776	12	BG323761	BG323761 602421865
36	242	15.6	846	13	BI116063	BI116063 602866459
37	241.2	15.5	1080	13	BM467429	BM467429 AGENCOURT
38	239.4	15.4	756	12	BG386139	BG386139 602455319
39	239	15.4	897	14	BQ727393	BQ727393 AGENCOURT
40	238.4	15.3	697	13	BI855975	BI855975 603383349
41	237.4	15.3	843	12	BE792381	BE792381 601585118
42	236.2	15.2	814	13	BI772615	BI772615 603057607
43	233.6	15.0	759	12	BF536680	BF536680 602048656
44	232.6	15.0	274	9	AA696993	AA696993 GM08866.5
45	229.8	14.8	668	14	BM849233	BM849233 K-EST0129

#### ALIGNMENTS

RESULT 1  
AI518473 790 bp mRNA linear EST 23-APR-2001  
LOCUS LD38001.5prime LD Drosophila melanogaster embryo P0R2 Drosophila  
DEFINITION melanogaster cDNA clone LD38001 5 similar to NIE: F8an0002863  
'signal transduction' located on: 2L 21C6-21C6;: 04/10/2001, mRNA  
sequence.  
ACCESSION AI518473 GI:13769352  
VERSION EST  
KEYWORDS fruit fly.  
SOURCE Drosophila melanogaster  
ORGANISM Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Eukaryota; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Phylogeny: Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE 1 (bases 1 to 790)  
AUTHORS Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,  
Lewis,S. and Rubin,G.M.  
TITLE BDGP/HMT Drosophila EST Project  
JOURNAL Unpublished (2001)  
COMMENT On Mar 16, 1999 this sequence version replaced gi:4424327.  
Other\_ESTs: LD38001.3prime  
Contact: Stapleton, M.  
BDGP  
Lawrence Berkeley National Lab  
One Cyclotron Rd, Berkeley, CA 94720, USA  
Fax: 510 486 6798  
Email: [http://www.fruitfly.org/EST\\_estefruitfly.berkeley.edu](http://www.fruitfly.org/EST_estefruitfly.berkeley.edu)  
hit genomic AE003589: arm:2L [301580,604052]



Db 321 CGCCACGACGAAGATGACGAGTTCATCGCCGGGACACGCCGAGGCTGTGTTTCGCTG 380

QY 364 AATTTTCAGCCCGGATGGTCTCATCTGCCAGTGGGAAGTGGCGACACACAGTGGCGATTG 423

Db 381 AATTTTCAGCCCGGATGGTCTCATCTGCCAGTGGGAAGTGGCGACACACAGTGGCGATTG 440

QY 424 TGGGATCTTAAACACAGACACCGCACTTCCTCTGCACAGGTCATACAGCAGTGGGTTCTG 483

Db 441 TGGGATCTTAAACACAGACACCGCACTTCCTCTGCACAGGTCATACAGCAGTGGGTTCTG 500

QY 484 TGGGATCTTAAACACAGACACCGCACTTCCTCTGCACAGGTCATACAGCAGTGGGTTCTG 543

Db 501 TGGGATCTTAAACACAGACACCGCACTTCCTCTGCACAGGTCATACAGCAGTGGGTTCTG 560

QY 544 ATCATCTGGGACCGGAGACCGGTCACAGAAAGGGGACCGCTTGCAGTGGGACCAAGAAA 603

Db 561 ATCATCTGGGACCGGAGACCGGTCACAGAAAGGGGACCGCTTGCAGTGGGACCAAGAAA 620

QY 604 CACATCAACTGGCTCGC 620

Db 621 CACATCAACTGGCTCGC 637

## RESULT 3

BF505998

LOCUS

DEFINITION

AT08344.5prime AT Drosophila melanogaster adult testes pOTB7  
Drosophila melanogaster cDNA clone AT08344 5 similar to Nle:  
FBan0002863 'signal transduction' located on: 2L 21C6-21C6;;  
04/07/2001, mRNA sequence.

ACCESSION

BF505998

VERSION

EST

SOURCE

fruit fly

ORGANISM

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 725)

REFERENCE

AUTHORS

Stapleton, M., Brokstein, P., Hong, L., Agbayani, A., Baxter, E., Berman

B., Carlson, J., Champe, C., Chavez, C., Chew, M., Dorsett, V., Farfan

D., Frise, E., George, R., Gonzalez, M., Guarin, H., Harris, N., Li, P.,

Liao, G., Miranda, A., Misra, S., Mungall, C.J., Nunoo, J., Pacleb, J.,

Paragas, V., Park, S., Phouanavong, S., Wan, K., Yu, C., Lewis, S.E.,

Ceiniker, S., and Rubin, G.M.

BDGP/HMI AT Drosophila EST Project

Unpublished (2000)

On Dec 6, 2000 this sequence version replaced gi:11589299.

Contact: Stapleton, M.

BDGP

Lawrence Berkeley National Lab

One Cyclotron Rd, Berkeley, CA 94720, USA

Fax: 510 486 6798

Email: [http://www.fruitfly.org/EST\\_estefruitfly.berkeley.edu](http://www.fruitfly.org/EST_estefruitfly.berkeley.edu)

hit genomic AB003589; arm:2L [301580,604052]

estimated-cyco:21B7-21C7; 04/07/2001

Plate: AT.83 row: D column: 8

High quality sequence stop: 661.

Location/Qualifiers

1..725

/organism="Drosophila melanogaster"

/db\_xref="taxon:7227"

/clone\_lib="AT08344"

/sex="male"

/dev\_stage="0-3 day old Ore-R males"

/lab\_host="Plates AT.10-AT.120: DH5-alpha. Plates

AT.121-AT.319: DH5-alpha Tona"

/note="Organ: ADULT testes; Vector: pOTB7; Site.1: EcoRI;

Site.2: XhoI. The mRNA for the testis library was made

from testes and seminal vesicles hand dissected from 0-3

day old Ore-R males. RNA kindly provided by the lab of

Margaret Fuller. Sized fractionated cDNAs were directly

BASE COUNT 187 a 182 c 215 g 141 t  
ORIGIN

Query Match 38.1%; Score 591.8; DB 12; Length 725;  
Best Local Similarity 99.2%; Pred. No. 1.5e-86;  
Matches 606; Conservative 0; Mismatches 2; Indels 3; Gaps 1;

QY 7 CAATAATGAGGAGGAGGACGAGGAGCAAGAGCGCCGACACATACATACAGGCGCGC 66

Db 115 CAGAAATTCGAGGAGAGCGGACACGAGCAAGAGCGCCGACACATACATACAGGCGCGC 174

QY 67 CTCGTTT---ACACGGCGAGGAAGCGCGCCGCCTGACCTGCGCGGAGGATCACT 123

Db 175 CTCGTTTTCGGACACGGCGAGGAAGCGCGCCGCCTGACCTGCGCGGAGGATCACT 234

QY 124 ACCCAGCAATTTGGGACTGATTTCACACGCGCTGCTGAAAAACGAGGAGGACCTCCATAT 183

Db 235 ACCCAGCAATTTGGGACTGATTTCACACGCGCTGCTGAAAAACGAGGAGGACCTCCATAT 294

QY 184 TTGTTTTTCGTCGGCGAGGATGAGATCAAGAGAGCCTTGAGGACACGTTGGACTTGGCG 243

Db 295 TTGTTTTTCGTCGGCGAGGATGAGATCAAGAGAGCCTTGAGGACACGTTGGACTTGGCG 354

QY 244 TCAGTGGACACCGAAAAACGTCGATATTGTGTATCAGCCACAGGCGGTTTCAAGTG 303

Db 355 TCAGTGGACACCGAAAAACGTCGATATTGTGTATCAGCCACAGGCGGTTTCAAGTG 414

QY 304 CGCCAGTGCACAGATGCACAGTTCATCGCCGGGACACGCGGAGGCTGTGGTTGCGTG 363

Db 415 CGCCAGTGCACAGATGCACAGTTCATCGCCGGGACACGCGGAGGCTGTGGTTGCGTG 474

QY 364 AATTTACGCCCGGATGGTGTCTATCTCGCCAGTGGAGTGGCGACACACAGTGGCGATTG 423

Db 475 AATTTACGCCCGGATGGTGTCTATCTCGCCAGTGGAGTGGCGACACACAGTGGCGATTG 534

QY 424 TGGGATCTTAAACACAGACACACCGCACTTCACCTGCACAGGTCATACAGCAGTGGTCTG 483

Db 535 TGGGATCTTAAACACAGACACACCGCACTTCACCTGCACAGGTCATACAGCAGTGGTCTG 594

QY 484 TGGGATCTTAAACACAGACACACCGCACTTCACCTGCACAGGTCATACAGCAGTGGTCTG 543

Db 595 TGGGATCTTAAACACAGACACACCGCACTTCACCTGCACAGGTCATACAGCAGTGGTCTG 654

QY 544 ATCATCTGGGACCGGAGACCGGTCAGCAGAGAGGCGGACCTTCAGTGGGACCAAGAAA 603

Db 655 ATCATCTGGGACCGGAGACCGGTCAGCAGAGAGGCGGACCTTCAGTGGGACCAAGAAA 714

QY 604 CACATCAACTG 614

Db 715 CACATCAACTG 725

## RESULT 4

BF505998

LOCUS

DEFINITION

SD26605.5prime SD Drosophila melanogaster Schneider L2 cell culture  
pOT2 Drosophila melanogaster cDNA clone SD26605 5 similar to Nle:  
FBan0002863 GO: [signal transduction (GO:0004871)] located on: 2L  
21C6-21C6;; 05/23/2001, mRNA sequence.

ACCESSION

BF505998

VERSION

EST

KEYWORDS

fruit fly

ORGANISM

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 622)

REFERENCE

AUTHORS

Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G.,

Lewis, S., and Rubin, G.M.

BDGP/HMI Drosophila EST Project

Unpublished (2001)





```
Db 277 TCAGTGACACCGAAGAGCGTGAATGATATGTTGATATCAGCCACAGCGCGTTTCAAAAGTG 336
Qy 304 CGCCCACTGACAAAGATGACAGAGTTCCATGCGGGACACCGCAGAGCTGTGGTTGCGCTG 363
Db 337 CGCCCACTGACAAAGATGACAGAGTTCCATGCGGGACACCGCAGAGCTGTGGTTGCGCTG 396
Qy 364 AATTTACAGCCCGGATGCTCATCTGCCAGTGGAGTGGCGACACACACAGTGCAGATTG 423
Db 397 AATTTACAGCCCGGATGCTCATCTGCCAGTGGAGTGGCGACACACAGTGCAGATTG 456
Qy 424 TGGGATCTTAACACAGACACCGCAGTTCACCTGACAGAGTCATAAGCAGTGGTTCG 483
Db 457 TGGGATCTTAACACAGACACCGCAGTTCACCTGACAGAGTCATAAGCAGTGGTTCG 516
Qy 484 TGGGATCTTAACACAGACACCGCAGTTCACCTGACAGAGTCATAAGCAGTGGTTCG 543
Db 517 TGGGATCTTAACACAGACACCGCAGTTCACCTGACAGAGTCATAAGCAGTGGTTCG 576
Qy 544 ATCATCTGGGACCCGGAGACGGGTGACGAG 573
Db 577 ATCATCTGGGACCCGGAGACGGGTGACGAG 606

RESULT 6
BI639652
LOCUS SD22291.5prime SD Drosophila melanogaster Schneider L2 cell culture EST 10-SEP-2001
DEFINITION pOT2 Drosophila melanogaster cDNA clone SD22291 5 similar to Nle:
FBan002863 GO:[signal transduction (GO:0004871)] located on: 2L
21C6-21C6; 05/19/2001, mRNA sequence.
ACCESSION BI639652
VERSION BI639652.1 GI:15541862
KEYWORDS EST.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 626)
Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
Lewis,S. and Rubin,G.M.
BDGP/HMI Drosophila EST Project
Unpublished (2001)
Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST. est@fruitfly.berkeley.edu
hit genomic AE003589; arm:2L [301580,604052]
estimated-cyto:21B7-21C7: 05/19/2001
Plate: SD.222 row: H column: 7
High quality sequence stop: 544.
Location/Qualifiers
1. 626
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="SD Drosophila melanogaster Schneider L2 cell
culture pOT2"
/lab_host="DH5-alpha"
/notes="Vector: pOT2; Site_1: EcoRI; Site_2: XhoI; Sized
fractionated cDNAs were directly ligated into pOT2.
Plasmid cDNA library."

BASE COUNT 155 a 158 c 188 g 124 t 1 others
ORIGIN
Query Match 35.3%; Score 548.8; DB 13; Length 626;
Best Local Similarity 98.9%; Pred. No. 1.5e-79;
Matches 563; Conservative 0; Mismatches 3; Indels 3; Gaps 1;

Qy 7 CAAAATATGACGAGACGGACACGAGCAAGCCACGACATACGATACAGCGCGC 66
|||||
```









Drive Rockville, Maryland 20850, USA Fax : (1) 301 610  
8371 Email : fliang@lifetech.com URL :  
http://fulllength.invitrogen.com

BASE COUNT 195 a 293 c 291 g 184 t 12 others  
ORIGIN

```
Query Match 21.3%; Score 331.6; DB 9; Length 975;
Best Local Similarity 60.2%; Pred. No. 1.7e-44;
Matches 532; Conservative 9; Mismatches 343; Indels 0; Gaps 0;

QY 54 GATACAGCGCGCTGCTTTACAGCGGAGAGAGCGCGCGCAATACGACCTGCGCGC 113
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 52 GCTAGTCAGTCCAGGATGAGCGGCGAGCTGCTGGGTCCCGTTCAGCTGCGCGT 111
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 114 AGGAATCACTACCCAGCAATTTGCGACGCTGATTCGACGCGCTGGAACAGGAGAC 173
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 112 GGACATCAACCCGAGAGCTGCGAGCTGCTGTCACAGCGCTTACGCGCAGGAGATCC 171
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 174 CACTCCATATTTGTTTCGTGGCGGAGGATGAGATCAAGAGAGCTTGGAGGACAGTT 233
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 172 CCGCCACTGGCTTCTTTKTCACGATGCTGAGATCTCTCTCACTGGGGAARACGTT 231
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 234 GGAATTCGGCTCAGTGGACACCGCAACGCTGATCGATATTGTATCAGCCACAGCGGT 293
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 232 GGAGTCCAGCGAGTGGAGACARAAAGTCTTACATCATCTACGAGCAGCGCTAT 291
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 294 TTTCAAAGTGGCGGAGTGAACAGTGCAGAGTTCATGCGCGGACACCGCGAGCTGT 353
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 292 CTTCARAGTCCGGCTGTGACTCGCTCCACAGCTCTTGGAGGTTCAGAGTGGCGAT 351
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 354 GGTTCGCTGATTTACGCGCGGATGTGCTATCTCGCGAGTGAAGTGGCGACACAC 413
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 352 CATTTCTGTGGCTTCAKCCCTACGGGAARATACCTTGGCCAGTGGCTCTGGAGACAC 411
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 414 AGTCGATTTGGGATTTAAACAGAGACACCGCACTTCACTGCGAGCTGATGAAGA 473
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 412 CGTGGCTTCTGGGATCTCAGCAGACAGACACCATTTTCAATGCAAGGACACARACA 471
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 474 GTGGTTCGTGCTGATCTCGGCTCCGAGTGCAGAGTTCGCGCGGAGCGGTTGCAAG 533
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 472 CTGGGCTTCTAGTATATCTGCTCTCCAGATGGCAGAGTGGCTTCAAGTGCAGAA 531
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 534 GGGCTCTATATCATCTGGGACCGGAGCGGTGACGAGAGGGGCGACCTTGAGTGG 593
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 532 TGGCGAGATTTCTCTGGGACCAACAGCAGGGAAGCAGGTGGGCGAGGACCTCGCTGG 591
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 594 GCACAAGAACACATCACTGCTCGCTGGGAACCGTATCATCGGATCCGAGTGCAG 653
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 592 CCACAGCAAGTGGATCAGAGGCTGAGCTGGGAGCGCTTCCATCGAACCCTGAGTGG 651
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 654 GAACCTTGCTCGCGCAGTGAGAGCGGAGTGCAGTGGGATTTGGGACGTAAATTTGG 713
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 652 CTATGTGGCCAGAGCTCCAGGATGGCAGTGGCGGATCTGGGACAACTCCAGGCGG 711
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 714 GTGCGTTATGACATTTGGCGGACACAAATGCTGTGACAGAGTGAAGTGGGTGGAG 773
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 712 CTGTGAGCGATCTCTCACCGGGCACACCCAGTGGTCACTGTCTCCGTTGGGAGGGGA 771
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 774 GGGCTTATTTATACATCTCCAAAGATCGCAGTGAAGATGTGGGCGACGCTGATGG 833
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 772 CGGGCTTCTACTCTGCTCTCCAGGACCGCACCACTCAAAATCTGGAGAGCTCATGCG 831
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 834 AATCTGTGCGGAGTCTCTGCGCAAGCTCACTGGGTAAACACATTCGCTGAGGAC 893
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 832 TGTGCTGTGCGGAGTCTGCAAGGCCACGCCACTGGGTGAACACATGTCCTCAGC 891
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 894 CGATAGCTCTGCGGACTGTGCTCAATTCATTCGCTGGAAGATC 937
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 892 TGACTATGCCCTGCGCACTGGGGCTTTGAAGTCTGAGGCTC 935
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 14  
AL526575

LOCUS AL526575 925 bp mRNA linear EST 13-FEB-2001  
DEFINITION AL526575 LTI\_NFL003\_NBC3 Homo sapiens cDNA clone CS0DC014YN14 5  
prime, mRNA sequence.  
ACCESSION AL526575  
VERSION AL526575.1 GI:12790068  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 925)  
AUTHORS Li, W.B., Gruber, C., Jesse, J., and Polayes, D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished (2001)  
COMMENT Contact: Genoscope  
Genoscope, Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES  
source

1..925  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="CS0DC014YN14"  
/clone.lib="LTI\_NFL003\_NBC3"  
/sex="male"  
/tissue.type="neuroblastoma cells"  
/lab.host="DH10B"  
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA  
was primed with a NotI-Oligo(dT) primer. Five prime end  
enriched, double-stranded cDNA was digested with Not I and  
cloned into the Not I and Eco RV sites of the pCMVSPORT 6  
vector. Library was normalized. Library was constructed  
by Life Technologies. Contact : Feng Liang Life  
Technologies, a division of Invitrogen 9800 Medical Center  
Drive Rockville, Maryland 20850, USA Fax : (1) 301 610  
8371 Email : fliang@lifetech.com URL :  
http://fulllength.invitrogen.com"

BASE COUNT 183 a 278 c 290 g 173 t 1 others  
ORIGIN

Query Match 20.2%; Score 314; DB 9; Length 925;  
Best Local Similarity 60.7%; Pred. No. 1.2e-41;  
Matches 529; Conservative 1; Mismatches 341; Indels 1; Gaps 1;

```
QY 54 GATACAGCGCGCTCGCTTTACACGGGCGAGGAAGCGCGCCGCAATGACCTGCGCGC 113
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 54 GCTAGTGCAGTTCCAGGATGAGGCGGCGAGCTGCTGGGTTCGCCGTTCCAGCTGCCGT 113
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 114 AGGAATCACTACCCAGCAATTTGGGACTGATTTGCAACGCGCTGCTGAAACAGGAGAC 173
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 114 GGACATCACCCCGGACAGGCTGCGAGCTGTGTGCAACGCGCTACTGGCCCGAGGATCC 173
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 174 CACTCCATATTTGTTTTCGTGGCGGAGGATGAGATCAAGAAGAGCCTGGAGGACAGCTT 233
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 174 CTGTCCACTGGCTTTCTTTGTCACGATGCTGAGATGCTCTCTCACTGGGGAAGACGTT 233
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 234 GGATTTGGCGTCAGTGGACACCGAAAACGCTGATCGATATTGTGTATCAGCCACAGCGCGT 293
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 234 GGATGCCAGGCGAGTGGAGACAGAGAAGGTCTCTAGACATCATCTTACCAGCCACAGCTAT 293
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 294 TTTCAAAGTGGCGCGCAGTCAAGATGCAGAGTTCCATGCCGGGACACCGCGAGGCTGT 353
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 294 CTTTCAGAGTCCGGGCTGTGACTCGTGCACAGCTCTTGGAGGGTCAAGTGGAGGAGTCC 353
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 354 GGTTCGCTGGAATTTTCAGCGCGGATGGTCTCATCTCCGAGTGAAGTGGAGGACACCCAC 413
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 354 CATTTCTGTGGCTTTCAGCCCTACGGGAAAGTACCTGGCCAGTGGCTCTGGAGACACCCAC 413
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 414 AGTGCAGTTTGGGATCTTAACACAGAGACACCGCACCTTCACTTGCACAGGTCATAAGCA 473
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 414 CGTGGCTTCTGGGATCTCAGCAGACAGACACCATTTTACATTCGAAGGACACAGACA 473
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 474 GTGGGTTCGTGCTGCTATCTCTCGGCTCCGAGTGGCAACGAGTTGGCCAGGCTTGCAAGC 533
```

Db 474 CTGGGCTCTTAGTATATCTCTGCTCCAGATGGCAAGAGCTGGCTCAGGCTCAAGAA 533  
 Qy 534 GGGCTCTATATCATCTGGGACCCGGAGACGGGTTCAGCAGAAAGGGCGACCTTTAGTGG 593  
 Db 534 TGGCAGATTCTCTCTGGGACCCCAAGCACAGGGAAGAGTGGGCGAGGACCTTCGCTG 593  
 Qy 594 GCACAGAAGAACATCACTGCTCGCTGGGACCGGTATCATCGATCCGGATCCGGAGTCAG 653  
 Db 594 CCACAGCAAGTGGATCACAGGCTGAGCTGGGAGCCCTCCATCGCAACCTCGATGCTCG 653  
 Qy 654 GAACTTGTCTTCGGCAGTGGAGACGGGGAGTCCCGGATTTGGGAGCTAAATTTGGGCCA 713  
 Db 654 CTATGTGGCCAGCAGCTCCAGGATGGCAGTGTGCGGATCTGGGACAACTTCGAGGCCG 713  
 Qy 714 GTGCTTATGAACATTTCCGGACACACAAATGCTGTGACAGCAGTGTGAGTGGGTGGAGC 773  
 Db 714 CTGTGAGCGATCTCTACCGGGCACACCCAGTCCGTGCTCTCGGTGGGAGGGGA 773  
 Qy 774 GGGCTTATTATACATCTCTCAAGATCGCACAGTGAAGATGTGGCGAGCAGCTGATGG 833  
 Db 774 CGGGCTTCTACTCTCTCTCCAGGACCGCACCATCAAGTCTGGAGAGCTCATGACGG 833  
 Qy 834 ATCTTGTGGCGGACGTTCTCTGGCCAGCTCACTGGGTAACACACATTTGGCTGAGCAC 893  
 Db 834 TGTGCTGGCGGAC-TCTGCAAGGACCGGCACTGGGTGAACACCATGGCCCTCAACAC 892  
 Qy 894 CGATTAGCTCTCGGCACATGGTCCATTCCATC 925  
 Db 893 TGACTATGCTCGGCACCTGGGGCCCTTTGAAC 924

## RESULT 15

AL530778 1114 bp mRNA linear EST 13-FEB-2001  
 LOCUS AL530778.LTI\_NFL001\_NBC4 Homo sapiens cDNA clone CS0DD008YA24 5  
 DEFINITION Prime mRNA sequence.

ACCESSION AL530778.1 GI:12794271  
 VERSION 1  
 KEYWORDS EST.

SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 1114)  
 AUTHORS Li, W.B., Gruber, C., Jessee, J., and Polayes, D.  
 TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished (2001)  
 COMMENT Contact: Genoscope

Genoscope - Centre National de Sequencage  
 BP 191 91006 EVRY cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.  
 Location/Qualifiers

## FEATURES

source  
 1..1114  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="CS0DD008YA24"  
 /clone\_lib="LTI\_NFL001\_NBC4"  
 /sex="male"  
 /tissue\_type="neuroblastoma cells"  
 /lab\_host="DH10B"  
 /note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 233 a 320 c 351 g 180 t 30 others  
 ORIGIN

Query Match 19.4%; Score 301.4; DB 9; Length 1114;  
 Best Local Similarity 57.3%; Pred. No. 1.2e-39;  
 Matches 582; Conservative 19; Mismatches 406; Indels 8; Gaps 4;  
 QY 425 GGGATCTTAACACAGACAGACACCGCACTTTCACCTGCACAGGTCTAAGCAGTGGGTCTGT 484  
 Db 2 GGGAKCGCAGCAGAGACACACATTTTCAGCGAGGACACAGACACTCTGGGCGCTTA 61  
 QY 485 GCGTATCTCTGGGCTCCGGATGGCAAGCGTTGGCCAGCTTGAAGAGCGGCTC-TATA 543  
 Db 62 GGAGATSCGGGGCTCCAGAGGCGAAGAGCGGGCTTCAGGCKGCRAGARTGCCACAGT 121  
 QY 544 ATCATCTGGGACCCCGAGAGCGGTTCAGCAGAAAGGGGACCCCTTCAGTGGGCGCAAGAAA 603  
 Db 122 CTCGCTGGGACCAAGCAGCGGAAGCGGAGGAGGAGGAGCCCTCGCKGGCCACAGCAAG 181  
 QY 604 CACATCAACTGCTCTCGCTGGGAACCGTATCATCGGGGATCCGGAGTCCGAGAACTTGTCT 663  
 Db 182 TGGATCACAGSSTGAGCGGGAGGCCCTCCATCGCAACCTGAGTGGCGGATGTGGCC 241  
 QY 664 TCCGCACTGGAGACGGGAGCTGCCGGATTTGGGAGCTAAATTTGGGCGAGTGCCTTATG 723  
 Db 242 AGCAGCTCCAAAGATGGCAGTGTGCGRKSXGGGACACAACTGCAGGCCGCTGTGAGCC 301  
 QY 724 AACATTCGGGACACACAAATCTGTGACAGCAGTGTGAGTGGGTGGAGCGGCGCTTATT 783  
 Db 302 ATCCCKACCGGGCACACCGAGTCGGTCACTGCTGAGTGGGAGGAGGAGGCGGCTTCTC 361  
 QY 784 TATACATCTCCAAAGATCGCACAGTGAAGATGTGGCGAGCAGCTGATGGAATCTTGTGC 843  
 Db 362 TACHCTGCTCCCGAGGACCGCACCATCAAGTCTGGAGAGCKATGACGGTGTGCKGTGC 421  
 QY 844 CGGAGCTTCTCTGGCAAGCTCACTGGGTAAACAACTTGCCTGAGCAGCAGTATAGTGC 903  
 Db 422 CGGACTCTGCAAGGCCACCGGCACTGGGTGAACACCATGGCCCTCAGCAGTACTATGCC 481  
 QY 904 CTGGGCACTGGTCCATTCCTCCGTGAAGATCGCTCCAGAGCC---ACCTCAGTTTG 960  
 Db 482 CTGCGCACTGGGCGCTTTGAACTCTCRGAGGCTCAGTGAATCCCAAGACCTCCAAGGA 541  
 QY 961 AGCATGAGGAATTCAGGAATCTGCCCTTGAAGCGCTTACAGCGCGTGTGCCCTGACGAG 1020  
 Db 542 TCCTTGAGGAGGTGAAGGAGAGGCTCTGAGCCGATACAACTCTGTGCGGGGCGCAGGGT 601  
 QY 1021 GTGGATCGCTGCTTCTCTGATGACAAACACCTCTATCTGTGG---CGGAACAAC 1077  
 Db 602 CCAGAGAGGCTGGTGTCTGCTCCGACGACTTCACCTTATCTGTGTGCTCCCGCAGCAG 661  
 QY 1078 CAGAACAGTGGCTTGGCGCATGACAGGCGACAGACGCTGTCAGAGTGTGAAATAT 1137  
 Db 662 GACAAAGGCTCTCACTCGGATGACAGGACACCAAGCTCTATCAACAGGCTCTTC 721  
 QY 1138 TGCGCGGATGAAGCTAATTCGCTCTCTTCAATTTGACAAGTCAAGTGTGCTGTGGCGA 1197  
 Db 722 TCTCTGACTCCCGCATCTGCTAGTGTAGTCCCTCTTTTTCACAACTCCATCAAGCTGGGAT 781  
 QY 1198 GCACGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1257  
 Db 782 GCAGAGAGGCGCAAGTACCTGCTTCCCTACGCGGCGCAGCTGGCTGCGGCTGTACAGATT 841  
 QY 1258 GCTGTGCTGGGAGCTCCGCTGATTTCTCGGAGCAAGACATCACTCACTCAAGATA 1317  
 Db 842 GCGTGGTCACTGACAGTGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 901  
 QY 1318 TGGAGTGTGACAGCAGGAAGAACTGCAAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1377  
 Db 902 TGGATGTGAAGGCCCAAGAGCTGGCCATGGACCTGCGGCGGCGGCGGAGGTATAT 961  
 QY 1378 GGAGTGAAGTGGGCGGCGGCTGTAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1432  
 Db 962 GCTGTGACTGGAGTCCAGATGGC-CAGAAATGCCAGTGTGGGAGGACAAATG 1015

Tue Jun 3 11:27:45 2003

us-09-830-980-2.rst

Page 12

Search completed: June 2, 2003, 06:52:31  
Job time : 2088 secs

• • •



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 2, 2003, 02:56:24 ; Search time 3985 Seconds  
(without alignments)  
11356.302 Million cell updates/sec

Title: US-09-830-980-2

Perfect score: 1555

Sequence: 1 aattcccaaaatgcagga.....aaaaaaaaaaaaaaaaaaaaa 1555

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl:\*

1: gb.ba.\*  
2: gb.htg.\*  
3: gb.in.\*  
4: gb.om.\*  
5: gb.ov.\*  
6: gb.pat.\*  
7: gb.ph.\*  
8: gb.pl.\*  
9: gb.pr.\*  
10: gb.ro.\*  
11: gb.sts.\*  
12: gb.sy.\*  
13: gb.un.\*  
14: gb.vi.\*  
15: em.ba.\*  
16: em.fun.\*  
17: em.in.\*  
18: em.in.\*  
19: em.mu.\*  
20: em.mu.\*  
21: em.or.\*  
22: em.ov.\*  
23: em.pat.\*  
24: em.ph.\*  
25: em.pl.\*  
26: em.ro.\*  
27: em.sts.\*  
28: em.un.\*  
29: em.vi.\*  
30: em.htg.hum.\*  
31: em.htg.inv.\*  
32: em.htg.other.\*  
33: em.htg.mus.\*  
34: em.htg.pln.\*  
35: em.htg.rod.\*  
36: em.htg.mam.\*  
37: em.htg.vrt.\*  
38: em.sy.\*  
39: em.htgo.hum.\*  
40: em.htgo.mus.\*  
41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match %	Length	DB	ID	Description
1	1555	100.0	1555	6	AX025192	Sequence
2	1528.2	98.3	1563	3	AY089286	Drosophila
3	1521	97.8	1521	3	DME012588	AY012588 Drosophila
c	1285.8	82.7	85095	3	AC004573	AC004573 Drosophila
5	1285.8	82.7	142257	2	AC017491	AC017491 Drosophila
6	1285.8	82.7	182726	3	AC008002	AC008002 Drosophila
c	1285.8	82.7	302473	3	AF003589	AE003589 Drosophila
8	478	30.7	1859	6	AX135811	AX135811 Sequence
9	477.4	30.7	1853	6	AK001320	AK001320 Homo sapi
10	477.4	30.7	1854	6	AK077627	AK077627 Sequence
11	477.4	30.7	1870	9	BC002884	BC002884 Homo sapi
12	477.4	30.7	2593	9	BC012075	BC012075 Homo sapi
13	477.4	30.7	2638	6	AX405715	AX405715 Sequence
14	476.6	30.6	1784	10	BC018399	BC018399 Mus muscu
15	466.2	30.0	1928	6	AX135807	AX135807 Sequence
16	464.6	29.9	2250	5	AF069737	AF069737 Xenopus l
17	249.4	16.0	1170	6	AX135809	AX135809 Sequence
c	210.8	13.6	316613	8	SCCHRII	X59720 S.cerevisia
18	210.8	13.6	316613	8	SCCHRII	AX488920 Sequence
19	159.8	10.3	1563	6	AX488920	AX488920 Sequence
20	120.4	7.7	378	6	AX072820	AX072820 Sequence
c	115.8	7.4	40305	8	SPCC18	AL031907 S.pombe c
22	111.6	7.2	1373	9	HSAJ5257	AJ005257 Homo sapi
23	90.4	5.8	31787	3	AF106576	AF106576 Caenorhab
c	87.8	5.6	215786	9	AC022916	AC022916 Homo sapi
24	86.2	5.5	203769	2	AL603745	AL603745 Mus muscu
25	86.2	5.5	216655	2	AL603745	AC120837 Mus muscu
26	86.2	5.5	216655	2	AL603745	AL713882 Mus muscu
27	84.6	5.4	132348	10	CNS07Y02	AP003588 Nostoc sp
c	82.8	5.3	334520	1	AP003588	AC096796 Rattus no
28	82.8	5.3	334520	1	AP003588	AC115671 Rattus no
c	79.8	5.1	34131	2	AC096796	BC026080 Homo sapi
30	79.8	5.1	178633	2	AC115671	AP003593 Nostoc sp
31	79.4	5.1	1844	9	BC026080	AF323585 Podospora
32	79.4	5.1	333550	1	AF003593	AF115313 Thermomon
33	78	5.0	4722	8	AF323585	AL832918 Homo sapi
34	77	5.0	4619	1	AF115313	AB009055 Arabidops
35	76.2	4.9	81575	8	AB009055	AY113628 Drosophila
c	74.4	4.8	2311	3	AY113628	AF323583 Podospora
37	74.4	4.8	2311	3	AY113628	AF100287 Trypanoso
38	73.6	4.7	5413	8	AF323583	AF323582 Podospora
39	72.8	4.7	1209	3	AF100287	AF323584 Podospora
40	72	4.6	5413	8	AF323582	AE012949 Chlorobiu
41	72	4.6	5413	8	AF323584	AL395015 T3 end of
c	72	4.6	10280	1	AE012949	L28125 Podospora a
42	70.6	4.5	5411	8	FLABTLP	AX430450 Sequence
43	70.6	4.5	5411	8	FLABTLP	
44	70.6	4.5	5411	8	FLABTLP	
45	68.4	4.4	2977	6	AX430450	

## ALIGNMENTS

RESULT 1  
AX025192  
LOCUS AX025192 Sequence 9 from Patent WO0026364.  
DEFINITION AX025192  
ACCESSION AX025192  
VERSION AX025192.1 GI:10186913  
KEYWORDS Drosophila sp.  
SOURCE Drosophila sp.  
ORGANISM Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 1555)  
REFERENCE Cohen,S., Bouwmeester,A. and Royet,J.  
AUTHORS Regulator of notch signaling activity  
TITLE





Db 1555 TGGGCTTATTAACAATCACTTGTACACGGTAAGAAAAATCTTAGGATAAAGTAA 1614  
QY 1504 AACCTCTCTAGTAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 1552  
Db 1615 AACCTCTCTAGTAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 1663

RESULT 3  
DME012588  
LOCUS Drosophila melanogaster mRNA for Notchless protein.  
DEFINITION  
ACCESSION AJ012588  
VERSION  
KEYWORDS notchless gene; WD40 repeats.  
SOURCE Drosophila melanogaster.  
ORGANISM Drosophila melanogaster.  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 1521)  
Royet, J.  
Direct Submission  
Submitted (28-Oct-1998) Royet J., Developmental Biology Programme,  
EMBL, Meyerhofstrasse 1 Heidelberg, 69117, GERMANY  
2 (bases 1 to 1521)  
Royet, J., Bouwmeester, T. and Cohen, S. M.  
Notchless encodes a novel WD40-repeat-containing protein that  
modulates Notch signalling activity  
EMBO J. 17 (24), 7351-7360 (1998)  
98077802  
9857191

Location/Qualifiers  
1. .1521  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
1. .1443  
/gene="notchless"  
1. .1443  
/gene="notchless"  
/note="WD40 repeat containing protein"  
/codon\_start=1  
/product="Notchless protein"  
/protein\_id="CAAI0070.1"  
/db\_xref="GI:4127781"  
/translation="MOETDEQEPHTIQARLVYTGEEAGPPIDLPAGITQQGLI  
CNALLKNEEATPYLFFVGEDEIKKLEDTLDIASVDENVIDIVYQPAVFKVRVTR  
CTSSMPGHAEEAVSLNFPDGHASGGDTVRLNDINTETPHFTCTGKOWKLV  
WAPDGKRLASGCKAGSIIWDPETGOOKRPLSGHKHINCLAWEPYHRDPCKRLAS  
ASGDGCRIDWVKGQCLMNIAGHTNNAVYAVRWGAGLIYTSKDRITVWRAADIL  
CRFTSGOAHVWNIALSTDYVLTGTFPHVVKDRSKSHLSLSELOESALKEVQAVCP  
DEVESLYSCSDNTLYLRNNONKCVBPWGHQVNDYKSPDKLIASAFDKSVR  
LWRASDQGYMATFRGHVQVYVYVANSRSLIVSGSKDSLKVMVQTKRLAQELPGH  
ADEVFGVDWAPDGSRAVSGKDKVTKLMAY"  
BASE COUNT 389 a 370 c 448 g 314 t  
ORIGIN

Query Match 97.8%; Score 1521; DB 3; Length 1521;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1521; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 ATCGAGGACGACACGAGCAGAGAGCCACCCACATACAGTACAGGCGCCCTCGTT 72  
Db 1 ATCGAGGACGACACGAGCAGAGAGCCACCCACATACAGTACAGGCGCCCTCGTT 60

QY 73 TACACGGGAGAGAGCGCGCCGCCCAATCGACCTGCGCGGAGGAATCACTACCCAGAA 132  
Db 61 TACACGGGAGAGAGCGCGCCGCCCAATCGACCTGCGCGGAGGAATCACTACCCAGAA 120

QY 133 TTGGGACTGATTGCAACGCGCTCTGAAAGAGAGGAGGACCTCCATATTTGTTTTTC 192  
Db 121 TTGGGACTGATTGCAACGCGCTCTGAAAGAGAGGAGGACCTCCATATTTGTTTTTC 180

QY 193 GTGGGCGAGATGAGATCAAGAGAGCCTTGAGGACACGTTGGACTTGGCGTCACTGGAG 252

181 GTGGCGAGGATGAGATCAGAGAGAGCCTGGAGGACACGTTGGAGCTTGGGCTCAGTGGAC 240  
QY 253 ACCGAAACGTCATCGATATTTGTATATCAGCCACAGGCGGTTTTCAAAGTGCAGCCAGTG 312  
Db 241 ACCGAAACGTCATCGATATTTGTATATCAGCCACAGGCGGTTTTCAAAGTGCAGCCAGTG 300  
QY 313 ACAAGATGCAGAGTTCCATGCCGGGACACGCCGAGGCTGTGGTTTCGTGAATTTTCAGC 372  
Db 301 ACAAGATGCAGAGTTCCATGCCGGGACACGCCGAGGCTGTGGTTTCGTGAATTTTCAGC 360  
QY 373 CCGGATGGTGCCTCATCTCCAGTGGAAAGTGGCGACACACAGTGCAGTATTTGGGATCTT 432  
Db 361 CCGGATGGTGCCTCATCTCCAGTGGAAAGTGGCGACACACAGTGCAGTATTTGGGATCTT 420  
QY 433 AACACAGAGACGACGCTCAGCTGCACAGGTCTAAGGAGTGGGTTCTGTGCTGATCC 492  
Db 421 AACACAGAGACGACGCTCAGCTGCACAGGTCTAAGGAGTGGGTTCTGTGCTGATCC 480  
QY 493 TGGGCTCCGGATGGCAACGGTTGGCCAGCGGTTGCAAGCGGGCTCTATAATCATCTGG 552  
Db 481 TGGGCTCCGGATGGCAACGGTTGGCCAGCGGTTGCAAGCGGGCTCTATAATCATCTGG 540  
QY 553 GACCCGAGACGCGGTCAGCAGAGAGGCGGACCCCTTGGTGGGCGACAAAGAACACATCAAC 612  
Db 541 GACCCGAGACGCGGTCAGCAGAGAGGCGGACCCCTTGGTGGGCGACAAAGAACACATCAAC 600  
QY 613 TGCTCGCTGGGACGATCATCTCGGATCCGGATCCGGATCCGGATCCGGATCCGGATCC 672  
Db 601 TGCTCGCTGGGACGATCATCTCGGATCCGGATCCGGATCCGGATCCGGATCCGGATCC 660  
QY 673 GGAGACGGGAGCTCCGGATTTGGGAGCTGAGTAAATTTGGGCGGCGCTTATTTATACATCC 732  
Db 661 GGAGACGGGAGCTCCGGATTTGGGAGCTGAGTAAATTTGGGCGGCGCTTATTTATACATCC 720  
QY 733 GGACACACAAATGCTGTGACAGCAGTGGGATGGGATGGGCGGCGCTTATTTATACATCC 792  
Db 721 GGACACACAAATGCTGTGACAGCAGTGGGATGGGATGGGCGGCGCTTATTTATACATCC 780  
QY 793 TCCAAAGATCGCAGAGTGAAGTGGCGAGCAGTGGATGAATTTGTCGGGAGGTC 852  
Db 781 TCCAAAGATCGCAGAGTGAAGTGGCGAGCAGTGGATGAATTTGTCGGGAGGTC 840  
QY 853 TCTGGCCAGCTCACTGGGTAACAACATTTGCGCTGAGCACCAGTACGCTCTGCGCACT 912  
Db 841 TCTGGCCAGCTCACTGGGTAACAACATTTGCGCTGAGCACCAGTACGCTCTGCGCACT 900  
QY 913 GTTCCATTCATCCCGTGAAGATCGCTCCAGAGCAGCTCAGTTTGGACTGAGGAA 972  
Db 901 GTTCCATTCATCCCGTGAAGATCGCTCCAGAGCAGCTCAGTTTGGACTGAGGAA 960  
QY 973 TTGCGAGGATTCGCTTGAACGCTACCGGCGGTGCGCTGACGAGGTGGAGTGGCTG 1032  
Db 961 TTGCGAGGATTCGCTTGAACGCTACCGGCGGTGCGCTGACGAGGTGGAGTGGCTG 1020  
QY 1033 GTTTCGCTGTCGATGACACACACCCCTCTATCTGTGCGGGAACAACAGAGTGCCTT 1092  
Db 1021 GTTTCGCTGTCGATGACACACACCCCTCTATCTGTGCGGGAACAACAGAGTGCCTT 1080  
QY 1093 GAGCGATCAGAGGACCAAGAGTGGTCAAGATGTGAATATTCGCGGATGTAAAG 1152  
Db 1081 GAGCGATCAGAGGACCAAGAGTGGTCAAGATGTGAATATTCGCGGATGTAAAG 1140  
QY 1153 CTAAATTCGCTGCTTCAATTTGACAGTCTGCTGTGGGAGGACGAGTGGTCTAG 1212  
Db 1141 CTAAATTCGCTGCTTCAATTTGACAGTCTGCTGTGGGAGGACGAGTGGTCTAG 1200  
QY 1213 TACATGGCCACCTTCGCGGCTCATGTGAGGCTGTGTTACGCGTGGCTGCTCGCGGAC 1272  
Db 1201 TACATGGCCACCTTCGCGGCTCATGTGAGGCTGTGTTACGCGTGGCTGCTCGCGGAC 1260  
QY 1273 TCCGCTTGATTTGTTCCGCGAGCAAGACTCACTTAAAGTATGGAGTGTGAGAGC 1332

Db	1261	TCGCGTCTGATTGTTTCGCGCAGCAAAAGACTCAACTCTAAAAGTATGAGTGTGCGAGC	1320
QY	1333	AAGAACTGCACAGAGTCTCGTGGACATCGGATGAGTGTTCGGAGTGGAGTGGCG	1392
Db	1321	AAGAACTGCACAGAGTCTCGTGGACATCGGATGAGTGTTCGGAGTGGAGTGGCG	1380
QY	1393	CCGATGGCTCTAGAGTGGCTCTGGTGGCAAGCAAAAGTTATAAGCTATGGGCTTAT	1452
Db	1381	CCGATGGCTCTAGAGTGGCTCTGGTGGCAAGCAAAAGTTATAAGCTATGGGCTTAT	1440
QY	1453	TAACAATCATTAAGTGTACACGTAAGAAATACTTAGGAATAAGTAAAGCTCCTG	1512
Db	1441	TAACAATCATTAAGTGTACACGTAAGAAATACTTAGGAATAAGTAAAGCTCCTG	1500
QY	1513	AGTAAAAAATAAAAAA 1533	
Db	1501	AGTAAAAAATAAAAAA 1521	
RESULT 4			
AC004573/c		85095 bp	linear
LOCUS		Drosophila melanogaster, chromosome 2L, region 21C5-21D1, pl clone	INV 30-JAN-1999
DEFINITION		DS07610, complete sequence.	
ACCESSION	AC004573		
VERSION	AC004573.1	GI:4204255	
KEYWORDS	HTG.		
SOURCE	Drosophila melanogaster.		
ORGANISM	Drosophila melanogaster		
REFERENCE			
AUTHORS	Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G., Butenhorf,C., Chamber,M., Chavez,C., Chew,M., Ciesiolka,L., Doyle,C.M., Farfan,D.E., Galle,R., Hummasti,S.R., Karra,K., Kearney,L., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L., Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P., Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Poon,L., Sequeleira,A., Sethi,H., Snir,E., Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and Rubin,G.M.		
TITLE	Sequencing of Drosophila chromosome 2L, region 21C5-21D1		
JOURNAL	Unpublished (1998)		
REFERENCE	2 (bases 1 to 85095)		
AUTHORS	Celniker,S.E., Agbayani,A., Galle,R., Svirskas,R.R., Hoskins,R.A., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G., Chavez,C., Chew,M., Doyle,C.M., Farfan,D.E., Flanagan,J., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L., Kim,S.H., Lee,B., Lomotan,M.A., Mak,J., Mazda,P., Mok,M.S., Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Snir,E., Twomey,B., Wan,K.H., Whitelaw,K.R., Yee,A., Zhang,R., Zieran,L.L. and Kimmel,B.E.		
TITLE	Direct Submission		
JOURNAL	Submitted (17-APR-1998) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA		
COMMENT	On Jan 30, 1999 this sequence version replaced gi:4164118. Sequence submitted by: Berkeley Drosophila Genome Project Lawrence Berkeley National Laboratory, MS 64-121 Berkeley, CA 94720		
FEATURES	source		
source	1..85095		
	/organism="Drosophila melanogaster"		
	/strain="v2: cn bw sp"		
	/db_xref="taxon:7227"		
	/chromosome="2L"		
	/map="21C5-21D1"		

BASE COUNT	22289	a	19671	c	20319	g	22816	t	
ORIGIN									
Query Match		82.7%		Score 1285.8;	DB 3;	Length 85095;			
Best Local Similarity		89.2%		Pred. No. 5.5e-311;					
Mismatches 1505;		Conservative		0;	Mismatches	7;	Indels	175;	Gaps
QY	7	CAAAAAATGAGGAGACGACACGAGGAGCAAGAGCCACCCACATACGATACAGGCGCG	66						
Db	27639	CAGAAAAATGAGGAGACGACACGAGGAGCAAGAGCCACCCACATACGATACAGGCGCG	27580						
QY	67	CTCGTTT---ACACGGCGGAGGAAGCGGCCGCCCAATCGACCTGCCGGCAGGAATCACT	123						
Db	27579	CTCGTTTCGGACACGCGGCGAGGAAGCGGCCGCCCAATCGACCTGCCGGCAGGAATCACT	27520						
QY	124	ACCCAGCAATTGGGACTGATTTGCAACGCGCTGCTGAAAAACGAGGAAGCACTCCCATAT	183						
Db	27519	ACCCAGCAATTGGGCTGATTTGCAACGCGCTGCTGAAAAACGAGGAAGCACTCCCATAT	27460						
QY	184	TTGTTTTCGTTGGCGGAGGATGAGATCAAGAGAGCCTGGAGGACACGTTGGACTTTGGCG	243						
Db	27459	TTGTTTTCGTTGGCGGAGGATGAGATCAAGAGAGCCTGGAGGACACGTTGGACTTTGGCG	27400						
QY	244	TCAGTGGACACCGAAAAACGTGATCGATATTGTATCATGCCACACAGCGGGTTCACAAAGTG	303						
Db	27399	TCAGTGGACACCGAAAAACGTGATCGATATTGTATCATGCCACACAGCGGGTTCACAAAGTG	27340						
QY	304	CGCCAGTGACAAAGATGACAGATTCCATGCCGGGACACCGCGAGGTGGTTTCGCTG	363						
Db	27339	CGCCAGTGACAAAGATGACAGATTCCATGCCGGGACACCGCGAGGTGGTTTCGCTG	27280						
QY	364	AATTTTCAGCCCGGATGTGCTCATCTCGCCAGTGGAGTGGCGACACACAGTGGCGATTG	423						
Db	27279	AATTTTCAGCCCGGATGTGCTCATCTCGCCAGTGGAGTGGCGACACACAGTGGCGATTG	27220						
QY	424	TGGATCTTAAACAGACACACCGCCTTCACCTGCACAGGTCTATACAGCTGGGTTCG	483						
Db	27219	TGGATCTTAAACAGACACACCGCCTTCACCTGCACAGGTCTATACAGCTGGGTTCG	27160						
QY	484	TGGTATCTCTGGCTCCCGATGGCAACGGTTGGCCAGCGGTGGCAAGCGGCTCATATA	543						
Db	27159	TGGTATCTCTGGCTCCCGATGGCAACGGTTGGCCAGCGGTGGCAAGCGGCTCATATA	27100						
QY	544	ATCATCTGGAGCCCGGAGACGGGTGACAGAGAGGGCGACCTTGGTGGCGACACAGAAA	603						
Db	27099	ATCATCTGGAGCCCGGAGACGGGTGACAGAGAGGGCGACCTTGGTGGCGACACAGAAA	27040						
QY	604	CACATCAACTGCTCGCTGGGAACCGTATCATCGCGATCGCGAGTGGAGAACTTGTCT	663						
Db	27039	CACATCAACTGCTCGCTGGGAACCGTATCATCGCGATCGCGAGTGGAGAACTTGTCT	26980						
QY	664	TCCGCGAGTGGAGACGGGAGTGGCGGATTTGGGAGCTAAAATTTGGGCGAGTGGCTTATG	723						
Db	26979	TCCGCGAGTGGAGACGGGAGTGGCGGATTTGGGAGCTAAAATTTGGGCGAGTGGCTTATG	26920						
QY	724	AACATTTCCCGGACACACAATGCTCTGACAGCAGTGGAGTGGGAGCGGCGCTTATT	783						
Db	26919	AACATTTCCCGGACACACAATGCTCTGACAGCAGTGGAGTGGGAGCGGCGCTTATT	26860						
QY	784	TATACATCTCTCCAAAGATGCGACAGTGAAGATGTGGGAGAGCTGATGTGAATCTTGTGC	843						
Db	26859	TATACATCTCTCCAAAGATGCGACAGTGAAGATGTGGGAGAGCTGATGTGAATCTTGTGC	26800						
QY	844	CGGAGCTCTCTGGCCCAAGCTCACCTGGTGGTAAACAACATTCGCTGAGACACGATTAAGTC	903						
Db	26799	CGGAGCTCTCTGGCCCAAGCTCACCTGGTGGTAAACAACATTCGCTGAGACACGATTAAGTC	26740						
QY	904	CTGCGCACTGGTCCATTCCATCCCGTGAAGATCGCTCCAAAGACCACTCAGTTTGAAC	963						
Db	26739	CTGCGCACTGGTCCATTCCATCCCGTGAAGATCGCTCCAAAGACCACTCAGTTTGAAC	26680						



Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H., Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C., Rogers,Y., An,H., Baldwin,D., Banazon,J., Beeson,K.Y., Busam,D.A., Carlson,J.K., Center,A., Champs,E., Davenport,L.B., Dietz,S.M., Dodson,K., Dorsett,V., Doull,L.E., Doyle,C., Dresnek,D., Farfan,D., Ferreira,S., Frise,E., Gale,R.F., Garg,N.S., George,R.A., Gonzalez,M., Houch,J., Hoskins,R.A., Hostin,D., Howland,T.J., Iqbal,S., Jalali,M., Kruse,D., Li,P., Mattai,B., Moshrefi,A., Mbatia,T.C., Moy,M., Murphy,B., Nelson,K.A., Nunoo,J., Pacleby,J., Fargas,V., Park,S., Patel,S., Pfeiffer,B., Scheeler,F., Shumanavong,S., Pittman,G.S., Puri,V., Richards,S., Williams,S.M., Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.

Sequencing of Drosophila chromosome 2L, region 21D-21E

Unpublished  
2 (bases 1 to 182726)

TITLE  
JOURNAL  
REFERENCE  
AUTHORS

Celniker,S.E., Agbayani,A., Arcalini,T.T., Baxter,E., Blazee,R.G., Butenhoff,C., Champey,M., Chavez,C., Chew,M., Ciesiolka,L., Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L., Kim,E., Lee,B., Lewis,S., Li,P., Lomutan,M.A., Mazda,P., Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleby,J.M., Park,S., Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E., Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and Rubin,G.M.

Direct Submission  
Submitted (09-JUL-1999) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA  
On Apr 6, 2001 this sequence version replaced gi:6532012.  
Sequence submitted by:  
Berkeley Drosophila Genome Project  
Lawrence Berkeley National Laboratory, MS 64-121  
Berkeley, CA 94720  
This sequence was assembled using end sequences from a whole genome shotgun and from subclones of this BAC and its neighboring clones. For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive Web site (<http://www.fruitfly.org/sequence/>) or send email to [bdgp@fruitfly.berkeley.edu](mailto:bdgp@fruitfly.berkeley.edu).

FEATURES  
source  
1. 182726  
/organism="Drosophila melanogaster"  
/strain="v; cn bw sp"  
/db\_xref="taxon:7227"  
/chromosome="2L"  
/map="21D-21E"  
/clone="BACR48E08 (D843)"  
/clone\_lib="RPCL-98 (Roswell Park Cancer Institute Drosophila melanogaster BAC library, partial EcoRI in pBac3.6)"  
BASE COUNT 49900 a 42515 c 41297 g 49014 t  
ORIGIN  
Query Match 82.7%; Score 1285.8; DB 3; Length 182726;  
Best Local Similarity 89.2%; Pred. No 6.1e-311;  
Matches 1505; Conservative 0; Mismatches 7; Indels 175; Gaps 4;  
  
QY 7 CAAAAATTCAGGAGCAGCACGAGCAAGAGGCCCGCCCAATCGACTGGCGGCAGGAATCACT 123  
Db 89258 CAGAAATTCAGGAGCAGCACGAGCAAGAGGCCCGCCCAATCGACTGGCGGCAGGAATCACT 89317  
QY 67 CTCCTTTT--ACAACGGGGGAGGAGCGCGCGCCCAATCGACTGGCGGCAGGAATCACT 123  
Db 89318 CTCCTTTTCGACACGCGGCGAGGAGCGCGCGCCCAATCGACTGGCGGCAGGAATCACT 89377  
QY 124 ACCCAGCAATTGGGACTGATTGGCAACGCGCTCTGTAAAACGAGGAGGCATCTCATAT 183  
Db 89378 ACCCAGCAATTGGGCTGATTGGCAACGCGCTCTGTAAAACGAGGAGGCATCTCATAT 89437  
QY 184 TTGTTTTTCGTGGCGGAGGATGAGATCAAGAAGAGCGCTTGAGGACACGTTGGACTTGGCG 243  
Db 89438 TTGTTTTTCGTGGCGGAGGATGAGATCAAGAAGAGCGCTTGAGGACACGTTGGACTTGGCG 89497





TITLE JOURNAL MEDLINE PUBMED 10731132 REFERENCE ADAMS AUTHORS TITLE JOURNAL COMMENT FEATURES source	Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R., Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wasserman,D.A., Weinstein,G.M., Weissenbach,J., Williams,S.M., Woodage,T., Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F., Zaveri,J.S., Zhan,M., Zhou,G., Zhao,Q., Zheng,L., Zheng,X.H., Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O., Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C. The genome sequence of <i>Drosophila melanogaster</i> Science 287 (5461), 2185-2195 (2000) 20196006	1. 302473 /organism="Drosophila melanogaster" /db_xref="taxon:7227" /chromosome="2L" /gene="P1c21C" /note="CG4574"	/map="21B7-21C1" /db_xref="FLYBASE:FBan0004574" /db_xref="FLYBASE:FBgn0004611" Join(3618..3958,11098..11376,11383..11475,19033..19179,27475..27612,28770..28903,29373..29558,29704..29901,30371..30553,31293..31747,32562..32796,34065..34210,35933..36146,36220..37083,37396..37843,37909..38301,38363..38507,38569..39613) /gene="P1c21C" /product="CT14752" /db_xref="FLYBASE:FBan0004574" Join(11383..11475,19033..19179,27475..27612,28770..28903,29373..29558,29704..29901,30371..30553,31293..31747,32562..32796,34065..34210,35933..36146,36220..37083,37396..37843,37909..38301,38363..38507,38569..39613) /gene="P1c21C" /note="P1c21C gene product" /codon_start=1 /protein_id="AAF51507.1" /db_xref="FLYBASE:FBan0004574" /db_xref="FLYBASE:FBgn0004611" /translation="MMSAGTYISTASVEVPQALQDGEKIRWDDSDSGTGTPTMRVD AKGFLYVQNNELDIDATIRVTRGQYAKRPDKNLQRIVTGLPQDTLEERTVT VCHGDFVMTFVNFCSTRDIAQLWTDLGLIKLAYSLAQLSGAIMFLQAKHTKLCLO VDKSRIPVKNILKLFQNKEDKRVKALDVTGLPSQKVDSTSVSKFQFEDFYNLK YLTQSEVERLFDSIVGNSKRCMSIAOLVFNKTDQPRNLNELYPYANPARAKEL IQQTEPNFNKQGLSLDGLFRLYLMGDDNPTMAPSKLDLDDMDQPMSPHYFINSNHL TYLTHGLTKGSVEIYRQCLLAGCKRVELDFWNGRTEEPVYVHGTFYVEIPEAKDVL EAIASAPKTSYEPVILSFENHCNPPQAKIANTCREIFGMDLDRPLDSHLEFNMD LPPPMARFKIITNKKHHHHHHHHKHPAQVGTFAANNKLTANSVDKAAQQQV LSHSDEGVTSTANGDVGATGTGSAAGTAGHAPPIQIROSXKSDSTGSSDSSS EDSEPLNTPNLPSGNEPPEKAQETAGAEISALVNYQPIHFSFENAEKKNRCY EMSSFDEKQATLLKRPTEFYNNKHOLSRVYPAGTRFDDSPFMQLFWNAGCOLVA LNFOPLDIAMOLNLGIFEYNAGNYLLKPEFMRSDRLDLPFAESTVDGIAGTTSIT VLSGOLFDRKANTFVEDVMYGLPADTVRKFRKTVTDNMGNPLDYDEEPFVKKVL PBLASIRIAYEEGKLIHGHVLPVLVIGLPCYRHNLRSEVGOPIALASIFLCVVVD YVPDLISNFAALANIKYSELEKRDQLSVLTDABEALGSADDELKSCQCKEIR PYESTLSPKRPSPISAAAMVDVTVDRDGGEDSISIVAPSIQHOSLDQSVST SRYVESQFQDVLVLAEPLEKLDHKSVEKREKKEKLESURKHDEKIKIAGQK SSPLGEGKPEATNKLKLSNKLNSCLSPSEPGVEIPACPLDLDGDSSESAADA GEDLAGSSSLDGRTOESRLRSACREYTSQYRELQEKYHEAIYSAAEKVLKTSQGT VRLKASLDKVTQEVYHOLQEARNEVKNLATVHRDRDLRMKREASSVYVREGAER VIKLQTFDRYDELQKHDSVRNALAEHRSKARQILDKAEARSCVSSNGFLVLRGP HHGCTGSGSSALSGNNLTNLIDAGAGSHSAISPAKSHNSIAAAEMKI" complement(11712..>12601) /gene="CG11912"	/map="21B7-21B8" /db_xref="FLYBASE:FBan0011912" /db_xref="FLYBASE:FBgn0031248" complement(11712..12601) /gene="CG11912" /product="CT37080" /db_xref="FLYBASE:FBan0011912" /db_xref="FLYBASE:FBgn0031248" complement(11773..12588) /gene="CG11912" /note="CG11912 gene product" /codon_start=1 /protein_id="AAF51509.1" /db_xref="GI:7296217" /db_xref="FLYBASE:FBan0011912" /db_xref="FLYBASE:FBgn0031248" /translation="MKQFAVIFALASVSALSVPQPGFPEGRIINGYEAKGEAPYI VSLQTSNSHFCAGLLDEVTIVTAACHLTYNQGOAVAGHSRTDOENVOIRKFTNAQ YVTHENYGGVGPNDIGLILKEEDAFDLNARDGSPVSAVSLPKGTQVSDGYL YWGRNDSGLLPNLQKLDALIVDYNECKAALPSNNSLAETNVCTHUPGRADGSCNGD SGGPLVSQSSRGAELIGIVSWGITPCLSTYTPSYVTSVSSFLPWIDENRKA" complement(13648..>14580) /gene="CG11911" /map="21B7-21B8" /db_xref="FLYBASE:FBan0011911" /db_xref="FLYBASE:FBgn0031249" complement(13648..14580) /gene="CG11911" /product="CT37076" /db_xref="FLYBASE:FBan0011911" /db_xref="FLYBASE:FBgn0031249" complement(13743..14576) /gene="CG11911" /note="CG11911 gene product" /codon_start=1 /protein_id="AAF51508.1" /db_xref="GI:7296216" /db_xref="FLYBASE:FBan0011911" /db_xref="FLYBASE:FBgn0031249" /translation="MKLTITVLVIALVAAQAGAKLSDKLKLVPFAFGFVINGTEAE PHAPYIVSLATNLYLKHSHICGGFLINKNDWIVTAACHISIEPVGMSIAGLHTRVDE YLQQRQVDFGRVHERYTGTVGVPDYDIALHVNESFIFENWVPATLPSREQVHEGHL YWGQPKSYIFSGAKTLQVTTQILNEECKELPESAPAESNICSSLSQSSKASCN F" complement(<49281..>51230) /gene="CG11907" /map="21C1-21C1" /db_xref="FLYBASE:FBan0011907" /db_xref="FLYBASE:FBgn0031250" complement(Join(49281..49459,49730..50200,50254..50554,50786..>51230)) /gene="CG11907" /product="CT33136" /db_xref="FLYBASE:FBan0011907" /db_xref="FLYBASE:FBgn0031250" complement(Join(49281..49459,49730..50200,50254..50554,50786..51049)) /gene="CG11907" /note="CG11907 gene product" /codon_start=1 /protein_id="AAF51506.1" /db_xref="GI:7296214" /db_xref="FLYBASE:FBan0011907" /db_xref="FLYBASE:FBgn0031250" /translation="MFGSDTFLVDALDEDETILGTGEGSFVAPRPODHHADQCLM ERQDEVLVSNPEPSGRLFYLYLIGICTGTPWNFVTAEDYWKYKFRNASINNT DLEELTPQKSTCDLALTATISCTTFLLNALFCHHSLFKMLGLTWLILFGV TGGTEINTDWOQFFLLITLIVVILNISAATMSGALGVAGLFPSEFTFAYVSGQA LGGLTALAPILVIAFDGTGNTAFIFFIVGVGLILICIVCYVILARPPFYRLEGG DKYKIVRAVSHNRNGSAREGLEPILRQVMSKIYLAISLALLYTLTSVYPAVTVL MOSYGHVWVAMVAFALNSNGYFTNILLIMAPKRVKQHEKELASSIMAAALSCGMAY GSLLSLVYVQML"	mrna
CDS					CDS
gene					gene
mrna					mrna
CDS					CDS
gene					gene



TITLE Novel genes tzap7/a, tzap7/b and tzap7 involved in t cell  
 JOURNAL activation and uses thereof  
 Patent: WO 0132614-A 5 10-MAY-2001;  
 Utuku, Nalan (DE)  
 FEATURES Location/Qualifiers  
 source 1..1859  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 CDS 20..1477  
 /note="unnamed protein product"  
 /codon\_start=1  
 /protein\_id="CAC39576.1"  
 /translation="MAAAVDAEAVARDVORLLVQFDEGGGLGSPFDPVVDITPDLR  
 /translation="MAAAVDAEAVARDVORLLVQFDEGGGLGSPFDPVVDITPDLR  
 QVRCNALLAQEDPLAFVFDHAEIVSLGKTESQAVETEKVLDIIYQPAIFVRRA  
 VTRCTSSLEGHSEAVISVAFSPGKYLASGGSDTTFVFDLSTETPHFTCKGHRHWL  
 STSPDGKLLASGCKNGOILLWDPSGTQVGRITLGHSGKLLYSLWEPHANPCRYD  
 VASSKDGSVRIWDTAGRCERILTGHTQSVTLRWGSDLLYSQDRTIKVRWPH  
 GVLCTQGHGHWNTMALSTDYALRTGAFPEAESYNPDLOGLQSLQELKRALSRYN  
 LYRGQPERLVSGSDDEFTLSPADKKPLTRMTGHOALINQYFSPSPSRIVASASE  
 DKSILMDRGRTKYLASLRGHVAAYTQIAWSADSRLLVSGSSDSTLKYVDVRAKLA  
 DLPGHADEVAVDWPDPQGVASGGKDLRIWR"  
 BASE COUNT 392 a 545 c 566 g 355 t 1 others  
 ORIGIN  
 Query Match 30.7%; Score 478; DB 6; Length 1859;  
 Best Local Similarity 59.8%; Pred. No. 8.5e-109;  
 Matches 837; Conservative 0; Mismatches 556; Indels 6; Gaps 2;  
 QY 54 GATPACAGCGCGCTCTGTTTACACGGCGGAGAGAACCGCGCGCGCAATTCGACCTGCGCGG 113  
 DB 70 GCTAGTGCAATGCCAGGATGAGCGCGGAGCTGTGTTCCCGCTTCGACGTGCGCGT 129  
 QY 114 AGCAATCACTACCCAGCAATTTGGGACTGATTTCGAACGCGCTGTGTAAGAAACAGGAAGC 173  
 DB 130 GGACATCACCCGAGCGTGCAGCTGTGTGCAACGCGCTACTTGGCCCGAGGAGTCC 189  
 QY 174 CACTCATTATTTGTTTTCGTCGCGGAGGATGAGATCAAGAGAGCGCTCGAGGACAGCTT 233  
 DB 190 CTTGCCACTTGGCTTTCTTTGTCACGATGTGTGAGATCGTCTCATCTGCGGGAAGAGCTT 249  
 QY 234 GGACTTGGCGCTCAGTGGACACCGAAACGCTGATCGATATTGTGTATCAGCCACACAGCGGT 293  
 DB 250 GGAGTCCCGAGCGTGGAGACAGAGAGGCTCTAGACATCATCTACACGACACAGGCTAT 309  
 QY 294 TTTCAAGTGCCTCAGTGCACAGATCCAGAGTTCCTCATGCGCGGACACCGCGAGCTGT 353  
 DB 310 CTTGAGTCCCGGCTGTGACTCGCTGCACCACTCTTTGGAGGCTCAGTGGAGGAGT 369  
 QY 354 GGTTCGCTGAATTTACGCGCGGATGGTGTCTCATCTCGCCAGTGGAGTGGCGACACCCAC 413  
 DB 370 CATTTCTGTGGCTTTCAGCCCTACGGGAAAGTACCTGGCCAGTGGCTCTGGAGACACCCAC 429  
 QY 414 AGTGGCATTTGGGATCTTAACACAGACACCGCACTTCACTTCACAGCTCATAGCA 473  
 DB 430 CGTGGCTTCTGGGATCTACACAGACACCAATTCATTCACAGGAGACACAGACA 489  
 QY 474 GTGGGTCTGTGGTATCCTGGCTCCGAGTGCACAAACGGTTGGCCAGCGGTTCGCAAGC 533  
 DB 490 CTGGGTCTTAGTATATCTGCTTCCAGATGGCAAGAAGCTGGCTCAGGCTGCAAGA 549  
 QY 534 GGCTCTAATCACTGGACCGGAGCGCTGACAGAGAGGCGGACCGCTTGAGTGG 593  
 DB 550 TGGCGAGATCTCTCTGGAGCCCAACACAGGAGGAGGTTGGCGAGACCTTCGCTGG 609  
 QY 594 GCACAGAAACATCAACTCGCTCGCTGGGAACCGTATCATCGCGATCGGAGTGCAG 653  
 DB 610 CCACAGCAAGTGGATCACAGGCGCTGAGTGGGAGCGCTCCATCGGAACCTTGAGTCCG 669  
 QY 654 GAACCTGCTTCGCGCAGTGGAGAGGAGTCCGCGATTTGGGAGCTAAATTTGGGCA 713  
 DB 670 CTATGTGGCCAGCAGCTCAAGGATGCGAGTGTGCGGATCTGGGACACAACTGCAGGCCG 729

QY 714 GTGCCATTATGACATTTGCCGGACACACAAATGCTGTGACAGCAGTGGATGGGTGGAGC 773  
 DB 730 CTGTGAGCGCATCTCTACCGGGCACACCCAGTGGTGGTCTGTCTCCGTGGGAGGGGA 789  
 QY 774 GGGCTTTATTTATACATCTCTCAAAAGATCGCACAGTGAAGATGTGGCGAGCAGCTGATGG 833  
 DB 790 CGGGCTTCTCTACTCTGTGCTCCAGGACCGCACCATCAAAAGTCTGGAGAGCTCATGACGG 849  
 QY 834 AATCTTGTGCGGACGCTTCTCTGCGCAAGCTCACTGGGTAAACAACATTCGCTCAGCAC 893  
 DB 850 TGTCTGTGCGGACTCTGCAAGGCCACGGCCACTGGTGAACACCATGGCTTCAGCAC 909  
 QY 894 CGATTAGCTCTCGGACTGTGCTCATTCATCCCGTGAAGATCGTCTCAAGGCC---A 950  
 DB 910 TGACTATGCCCTCGGACTTGGGGCTTTTGAACCTGCTGAGGCGCTCAGTTAATCCCAAGA 969  
 QY 951 CCTCAGTTTGAAGCACTGAGGAATTCAGGAATCTCCCTTGAAGCGCTTACCAAGCGCTGTG 1010  
 DB 970 CTTCCAAAGGATCTTTCAGAGGAGTGAAGGAGAGGGCTCTGAGCGGATACAACCTCGTGG 1029  
 QY 1011 CCCTGACAGAGTGGAGTCTGCTGTTTCTCTGTCGGATGACACACCTCTATCTGTGG-- 1068  
 DB 1030 GGGCCAGGGTCCAGAGAGGCTGTGTCTGGCTCCGAGACTTCACTTATCTGTGTGTC 1089  
 QY 1069 -CGGAACAACACAGCAAGTGGCTTGAAGCGCATGACAGGCGCACAGACCTGCTCAACGA 1127  
 DB 1090 CCCAGCAGAGGACAAAAGAGCTCTCACTCGGATGACAGGACACAAAGCTCTCATCAACCA 1149  
 QY 1128 TGTGAATATTCGCCGATGTAAGCTAATTTGCGTCTGCTTCAATTTGACAAGTCAAGTCA 1187  
 DB 1150 GGTGNTCTTCTCTCTGACTCCCGGATCGTGGCTAGTGGCTCTCTTGAAGTTCATCA 1209  
 QY 1188 TCTGTGGCGAGCGAGCGATGGTCACTAGTACATGCGCCACCTTCCGGGGTCTGTCAGGCTGT 1247  
 DB 1210 GCTGTGGGATGGCAGGACGGCAAGTACCTGCTTCCCTACGCGGCGAGTGGTGGCGGT 1269  
 QY 1248 TTACACGCTTGGCTGTGCGGACTCCCGCTTGTATTTTTCGGCGAGCAAGAGACTCAAC 1307  
 DB 1270 GTACAGATGGCTGTGTCAGTGCAGTGCCTGCTGTCAGCGGCGAGCAGTCAAGCAC 1329  
 QY 1308 TCTAAAGTATGGAGTGTGCAGACAGAAATGTCACAGAGAGTGCCTGGAGATGCGGA 1367  
 DB 1330 ACTGAAGTGTGGGATGTGAAGCCCAAGAAAGTGGCCATGCTCCCGGCGACCGCGGA 1389  
 QY 1368 TGAGTGTTCGAGTGGAGTGGCGCCGATGGCTCTAGAGTTCGCTGTGGGCAAGGA 1427  
 DB 1390 TGAGTATATCTGTTGACTGGAGTCCAGATGCGCAGAGAGTGGCAAGTGGTGGGAAGA 1449  
 QY 1428 CAAGTTATAAGTATG 1446  
 DB 1450 CAATGCTCGCGGATATGG 1468

## RESULT 9

AK001320  
 LOCUS Homo sapiens cDNA FLJ10458 fis, clone NT2RP1001457, highly similar  
 DEFINITION to Homo sapiens partial mRNA for beta-transducin family protein.

AK001320  
 ACCSSION AK001320.1 Gi:7022501

VERSION oligo capping; fis (full insert sequence).

KEYWORDS Homo sapiens teratocarcinoma cell\_line:NT2 cDNA to mRNA,

SOURCE clone.lib:NT2RP1 clone:NT2RP1001457.

## ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

## REFERENCE

## AUTHORS

1  
 Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,  
 Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H.,  
 Nagatsuma,M., Hosoi,T., Kaku,Y., Kodaira,H., Kondo,H.,  
 Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K.,  
 Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K.,  
 Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A.,

Nakamura, Y., Nagahari, K., Masuho, Y., Ninomiya, K. and Iwayanagi, T.  
NEDO human cDNA sequencing project

BASE COUNT 378 a 549 c 565 g 361 t  
ORIGIN  
DLFGRADEVIAVDSFGQQRVASGGKDKCLRIWRR\*

Query Match	30.7%	Score 477.4;	DB 9;	Length 1853;
Best Local Similarity	59.8%	Pred. No. 1.2e-108;		
Matches 837; Conservative	0;	Mismatches 556;	Indels 6;	Gaps 2;

Db	440	CGTGGCGTTCTGGGATCTCAGCACAGAGACACACATTTTCATATCAAGAGGACACAGACA	499
QY	474	GTGGGTTCTGTGGTATCCTGGGCTCCGGATGCGAAGCGTTTGGCCAGCGGTTGCAAGC	533
Db	500	CTGGGTCCTTAGTATATCTGTCTCCAGATGCGAAGAAGCTGGGCTCAGGCTGCAAGAA	559
QY	534	GGGCTCTAATAATCTCTGGGACCCGGAGACGGGTTCAGCAGAAAGGGGCGACCTTTGAGTGG	593
Db	560	TGGCCAGATCTTCCTCTGGGACCCCAAGCACAGGGAAGCAGGTGGGCGAGGACCTTCGCTGG	619
QY	594	GCAAGAAGAACATCAACTCCCTGCGCTGGGAACCGTATCATCGCGATCCGGAGTGCAG	653
Db	620	CCACACCAAGTGGATCACAGGCCCTGAGCTGGGAGCCCTCCATCGGAACCTGAGTCCGG	679
QY	654	GAACCTTGTCCGCCAGTGGAGACGGGGACTCCCGGATTTGGGACGTAAATTTGGGGCCA	713
Db	680	CTATGTGGCCAGACGCTCCAAGAGTAGGCACTGTGCGGATCTGGGACACAATCGCAGCCG	739
QY	714	GTGCCCTTATGAACATTTGCCGGACACACAATGCTGTGACAGCAGTGAGATGGGTGGAGC	773
Db	740	CTGTGAGCGCATCCTCACCGGGCACACCCAGTGGGTCACTGTCTCCGGTGGGAGGGGA	799
QY	774	GGGCTCTATTATACATCTCTCAAGATCGCACAGTAGAGATGTGGCGAGCAGCTGATGG	833
Db	800	CGGGCTCTCTACTCTGCTCTCCAGGACCGCACCATCAAAAGTCTGGAGAGCTCATGACGG	859
QY	834	AACTFTGTGCGGAGTTCTCTGGCCAAAGTCACCTGGGTAACACACATTCGCTGAGCAC	893
Db	860	TGTGCTGTGCGGAGCTGTGAAGGCCACGCCCACTGGGTGAACACCATGTCCTCAGCAC	919
QY	894	CGATTACGTCCTCGGCACATGGTCTCATTCGGTGAAGGATCGCTCCAAGAGCC---A	950
Db	920	TGACTATGCCCTGCGCACTGGGCCCCTTGAACCTGCTGAGGCCCTCAGTTATCCCCAGA	979
QY	951	CCTCAGCTTTGAGCACTGAGGAATTCAGAGATCTGCTTTGAAGCGCTACAGAGCGCTGGT	1010
Db	980	CCTCCAAGATCTCTTGCAGAGTTGAAGGAGAGGGCTCTGAGCCGATACAACTCCTGGC	1039
QY	1011	CCCTGACGAGGTGAGTCGCTGGTTCTCTGTTTCGGATGACAAACACCTCTATCTGTGG--	1068
Db	1040	GGGCCACAGGCTCCAGAGAGGCTGTGTCTGGCTCCGACGACTTCACCTTATTCCTGTGGTC	1099
QY	1069	-CGGAACAACAGAACAAAGTTCGCTTGAGCGCATGACAGGGCACAGAACGTGTCAACGA	1127
Db	1100	CCGACGAGGACAAAGAAGCTCTCACTCGGATGACAGGACACCAAGCTCTCATCAACCA	1159
QY	1128	TGTGAATATTCGCGGAGTGAAGCTAATTCGCTCTGCTTCATTTGACAAGTCAGTGGG	1187
Db	1160	GGTGTCTCTCTCTCTGACTCCGCGATCGTGCTAGTGCCTCTTTGACAAGTCCATCAA	1219
QY	1188	TCTGTGGGAGCCAGCATGGTCAGTACATGCCCACCTTCGCGGGTCTATGTGAGGCTGT	1247
Db	1220	GCTGTGGGATGGCAGGACGGGAAGTACCTGGCTTCCCTACGCGGCCACGTGGCTGCCGT	1279
QY	1248	TTTACAGGTTGCCGTGTCGCGGACTCCCGCTTGAATTTTCCGGGAGCAAGAGCTCAAC	1307
Db	1280	GTACCAGATTCGGTGGTCACTGACAGTCGGCTCTGCTCAGCGCGCAGCTGACAGCAC	1339
QY	1308	CTAAAGATATGGAGTTCGAGACGAGAAGAACTGGCACAGAGCTGCTGGACATCGGA	1367
Db	1340	ACTGAAGGTGTGGGATGTGAAGCCCCAGAAGCTGGCCATGGACCTGCCCGCCACGCGGA	1399
QY	1368	TCAGGTGTTCGGAGTGCAGTGGCGCCCGATGGCTCTAGAGTTGCCCTCTGGTGGGCAAGA	1427
Db	1400	TCAGGTATATCTCTTGACTGGAGTCCAGATGGGOCAGAGATGGTGAAGTGGTGGGAAGA	1459
QY	1428	CAAGGTTATAAGCTATGG	1446
Db	1460	CAATGCTCCGGATATGG	1478

RESULT 10

AX077627  
LOCUS AX077627 1854 bp DNA linear PAT 22-FEB-2001  
DEFINITION Sequence 82 from Patent WO0105970.  
ACCESSION AX077627  
VERSION AX077627.1 GI:13122009  
KEYWORDS human.  
SOURCE  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 1854)  
AUTHORS Yue, H., Tang, Y. T., Bandman, O., Hillman, J. L., Lal, P., Au-Young, J.,  
Reddy, R., Yang, J., Baughn, M. R., Lu, D. A., Azimzai, Y. and  
Patterson, C.  
TITLE Gtp-binding protein associated factors  
JOURNAL Patent: WO 0105970-A 82 25-JAN-2001;  
Incyte Genomics, Inc. (US)  
FEATURES  
source  
1..1854  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/note="Incyte ID No: 2216640CB1"  
BASE COUNT 389 a 548 c 563 g 354 t  
ORIGIN  
Query Match 30.7%; Score 477.4; DB 6; Length 1854;  
Best Local Similarity 59.8%; Pred. No. 1.2e-108;  
Matches 837; Conservative 0; Mismatches 556; Indels 6; Gaps 2;  
QY 54 GATACAGGCGCGCTCTGTTTACACGGCGGAGGAAGCGCGCGCAATCGACCTTCGCGCG 113  
DB 69 GCTAGTGCATCCAGGATGAGGGCGGCGAGCTGCTGGCTTCCCGCTTCGAGGTGCCGT 128  
QY 114 AGAATCACTACCGAATTTGGGACATTTCCAGCGGCTGTCTGAATAACAGGAGAC 173  
DB 129 GGACATCAACCGGACAGGCTGCAGCTCGTGTGCACGCGCTACTTGGCCAGGAGATCC 188  
QY 174 CACTCCATATTTTTCGTGGCGAGGATGAGATCAAGAAGAGCTGGAGGACAGCTT 233  
DB 189 CTTGCCACTGGCTTCTTTTGTCCAGATGCTGAGATCGTCTCTCACTTGGGGAAGCTT 248  
QY 234 GGACTTGGGCTCAGTGACACCGAAACGCTGATCGATATTTGTTATCAGCCACAGCGGCT 293  
DB 249 GGAGTCCAGGACGTGGACAGAGAGGCTCTAGACATCATCTACACACACAGGCTAT 308  
QY 294 TTTCAAGTGGCGCCAGTCAAGATGCAGAGTTCATGCCGGGACACCGCGAGGCTGT 353  
DB 309 CTTCCAGTCCCGGCTGTGACTCGCTGCACAGCTCCTTTGGAGGGCTCAGTGAGGCACT 368  
QY 354 GTTTTCGCTGAATTTACGCCCGGATGGTGTCTCATCTCGCCAGTGGAAAGTGGCGACACAC 413  
DB 369 CATTCTGTGGCTTCAGCCCTACGGGAAGTACCTGGCCAGTGGCTCTGGAGACACAC 428  
QY 414 AGTCCGATTTGGGATCTTACACAGACACCGCACTTCACTCGACAGGTCTAAGCA 473  
DB 429 CGTGCCTCTGGGATCTCAGCAGACAGACACACATTTACATGCAAGGGACACAGACA 488  
QY 474 GTGGCTTCTGCGTATCTCGGCTCCGGATGGCAACGCTTGGCCAGCGTTCGCAAGC 533  
DB 489 CTGGTCTCTAGTATATCTGGTCTCCAGATGGCAAGAGCTGGCCCTCAGGCTCGAAGAA 548  
QY 534 GGGCTCTATATCATCTGGGACCGCGAGACGGGTGACGAGAGGGGCGGACCTTTGAGTGG 593  
DB 549 TGGCAGATTTCTCTCTGGGACCCAAACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 608  
QY 594 GCACAGAAACACATCACTCGCTCGCTGGGACCGATATCATCGCGATTCGGAGTGCAG 653  
DB 609 CCACAGCAAGTGGATCAGAGGCTGAGCTGGGAGCCCTCCATGCGAACCCTGAGTGGCG 668  
QY 654 GAAACTTGTCTTCGCGCAGTGGAGGAGGAGGAGTGGCGGAGTGGAGGATTTGAATTTGGCGCA 713  
DB 669 CTATGTGGCCAGCAGCTCCCAAGGATGGCAGTGTGCGGATCTGGGACACACAACTGCGAGCGCG 728

QY 714 GTGCTTATGAACATTCGCGGACACACAAATGCTGTGACAGCACTGAGATGGGTGGAGC 773  
DB 729 CTGTGAGCGATCTCTCACCGGACACACCCAGCTCGGTACCTGTCTCCGGTGGGAGGGGA 788  
QY 774 GGGCTTATTTATACATCTCTCCAAAGATCGCACAGTGAAGATGTGGCGAGCAGCTGATGG 833  
DB 789 CGGCTTCTCTACTCTGCTGCCAGGACCGCACCATCAAAAGTCTGGAGAGCTCATGACGG 848  
QY 834 ATCTTGTGGCGGACGTTCTCTGCGGCAAGCTCACTGGTAAACAACTTCGCTGAGCAGC 893  
DB 849 TGTGCTGTGGCGGACGTTCTGCAAGGCGACCGGCTGGTGAACACCATTCGCTGAGCAGC 908  
QY 894 CGATTAGCTCTCTGCGCACTGTGCTCATTCATCCCGTGAAGAGTGTGCTCAAGAGCC--A 950  
DB 909 TGACTATGCTCTGCGCACTGTGGGCGCTTTGAACCTGCTGAGGCGCTCAGTTAATCCCAAGA 968  
QY 951 CCTCAGTTTGAGCACTGAGGAATTCAGGAATTCGCTTGAAGCGCTTACCAAGCGCTGTG 1010  
DB 969 CCTCAGGATCTCTGCAAGGATTTGAAGAGGAGGCTCTGAGCGGATACACCTCTGTGG 1028  
QY 1011 CCTGACGAGTGGAGTGTGCTGCTTCTGTTTGGATGACAAACACCTCTATCTGTGG-- 1068  
DB 1029 GGGCCAGGCTCCAGAGAGGCTGTGCTGCTGCTCCGAGCTTCACTTATCTGTGGTCT 1088  
QY 1069 -CGGAACAACCAAGCAAGTGTGCTGAGCGCATGACAGGCGACCAAGAGCTGTGCTCAACA 1127  
DB 1089 CCAGCAGAGGACAAAGGCTCTCACTCGGATGACAGGACCAAGCTCTCATCAACA 1148  
QY 1128 TGTGAATATTTCGCGGATGTAAGCTAATTCGCTGTCTTTCATTTGACAGTCAAGTCACT 1187  
DB 1149 GGTGCTCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1208  
QY 1188 TCTGCGGAGCGAGCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1247  
DB 1209 GCTGCGGATGCGAGGAGCGGCAAGTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1268  
QY 1248 TTACAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1307  
DB 1269 GTACAGATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1328  
QY 1308 TCTAAAGATGAGTGTGAGTGTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1367  
DB 1329 ACTGAAGTGTGGATGTGAAGGCGCCAGAGGCTGCGCATGCTGCTGCTGCTGCTGCTGCT 1388  
QY 1368 TGAGTGTTCGAGTGTGAGTGTGAGGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1427  
DB 1389 TGAGTGTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1448  
QY 1428 CAAAGTTATAAAGCTATGG 1446  
DB 1449 CAAATGCTCGGATATGG 1467  
RESULT 11  
BC002884  
LOCUS  
DEFINITION Homo sapiens, hypothetical protein similar to beta-transducin family, clone IMAGE:3943492, mRNA, partial cds.  
ACCESSION BC002884  
VERSION BC002884.1 GI:12804062  
KEYWORDS  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 1870)  
AUTHORS Strausberg, R.  
TITLE Direct Submission  
JOURNAL Submitted (05-FEB-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA  
REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>



	BASE COUNT	582 a	734 c	749 g	528 t	
	ORIGIN					
		Query Match	30.7%; Score 477.4; DB 9; Length 2593;			
		Best Local Similarity	59.8%; Pred. No. 1.3e-108;			
		Matches 837; Conservative	0; Mismatches 556; Indels 6; Gaps 2;			
QY	1428 CAAAGTTATTAAGCTATGG 1446					
Db	1438 CAATGCTCCGGATATGG 1456					
RESULT 12						
BCOL12075	2593 bp mRNA linear PRI 06-AUG-2001					
LOCUS	Homo sapiens, hypothetical protein similar to beta-transducin					
DEFINITION	family, clone MGC:19970 IMAGE:4652713, mRNA, complete cds.					
ACCESSION	BCOL12075					
VERSION	BCOL12075.1 GI:15082334					
KEYWORDS	MGC.					
SOURCE	Homo sapiens.					
ORGANISM	Homo sapiens					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
AUTHORS	Mammalia; Eutheraia; Primates; Catarrhini; Hominidae; Homo.					
TITLE	1 (bases 1 to 2593)					
JOURNAL	Direct Submission					
REMARK	Submitted (02-AUG-2001) National Institutes of Health, Mammalian					
COMMENT	Gene Collection (MGC), Cancer Genomics Office, National Cancer					
	Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,					
	USA					
	NIH-MGC Project URL: http://mgc.nci.nih.gov					
	Contact: MGC help desk					
	Email: qcqpb@femail.nih.gov					
	Tissue Procurement: ATCC					
	CDNA Library Preparation: Rubin Laboratory					
	CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)					
	DNA Sequencing by: Genome Sequence Centre,					
	BC Cancer Agency, Vancouver, BC, Canada					
	info@bcsc.bc.ca					
	Susanna Chan, Jennifer Asano, Ian Bosdet, Yaron Butterfield,					
	Stevana Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,					
	Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo					
	Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven					
	Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline					
	Schein, Duane Smalls, Michael Smith, Lorraine Spence, Jeff Stott,					
	Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,					
	George Yang, Scott Zuyderduyn, Marco Marra.					
	Clone distribution: MGC clone distribution information can be found					
	through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov					
	Series: IRAL Plate: 29 Row: o Column: 21					
	This clone was selected for full length sequencing because it					
	passed the following selection criteria: matched mRNA gi: 7022501.					
FEATURES	Location/Qualifiers					
source	1..2593					
	/organism="Homo sapiens"					
	/db_xref="LocustID:54475"					
	/db_xref="taxon:9606"					
	/clone="MGC:19970 IMAGE:4652713"					
	/tissue_type="Placenta, choriocarcinoma"					
	/clone_lib="NIH_MGC_21"					
	/lab_host="DH10B-R"					
	/note="Vector: pOTB7"					
	14..1471					
CDS	/codon_start=1					
	/product="hypothetical protein similar to beta-transducin					
	family"					
	/protein_id="AAH12075.1"					
	/db_xref="GI:15082335"					
	/translation="-MAAAVADEAVARDVQRLLVFODEGQLLGSPFDPVPDIIPDLR					
	OLVNALLAQEDPLPIAFVFDHAETIVSLKLTLESQAVTEKVFLDIIITQPQRIERVR					
	VTRCSSLSEGHSAVISVAFSPSTKYLSASGDFTVPWDLSTEPHFETCKGRHWRL					
	SISWSPDGKVLAKSGNQIILLWPSTGKQVSTGLAGHSKWITGSWEPLHANPECHY					
	VASSSKDGVIRIWDTAGRCERILTGHGTQVTCILRWGSDGLLYSASODRTIKVVRHD					
	GVLCTRLOQHGWNVNMTALSYDALTGAFPAEASVNPQDLQYSQELKEKRALSRYN					
	LVRGSCPERLVSSDDTFLEWSAEDEKKPLTRMTGNQALINVLFSPDSPRIVASAF					
	DKSLKWDGTGYLKLRSRGHVAAYOYLANSADRSLLVSSSDSTLKVDWKQAQKIAM					
	DYEGHADVYVNDPQGVNSGCCPKCIIEPP."					



Db 1024 GGGCCAGGGTCCAGAGAGGCTGGTGTCTGGCTCCGACGACTTACCTTATTCCTGTGGTC 1083  
QY -CGGAACAACAGAACAAAGTGGCTTGGAGCGCATGACAGGCGACACAGAGCTGGTCAACGA 1127  
Db 1084 CCACGACAGACACAAAGGCTCTCACTCGGATGACAGACACCAAGCTCTCATCAACCA 1143  
QY TGTGAATATTGCGCGGATGATAAGCTAATTCGCTGCTTTCATTTGACAAGTCAAGTCG 1187  
Db 1144 GGTGCTCTTCTCTCTGACTCCGCGATCGTGGCTAGTGGCTCTTTGACAAGTCCATCAA 1203  
QY TCTGTGGCGAGCCAGCATGGTTCAGTACATGCCACCTTCGCGGTCATGTCAGGCTGT 1247  
Db 1204 GCTGTGGGATGACAGGACGGCAAGTACCTGGCTTCCCTACGCGGCACAGTGGCTGCCGT 1263  
QY TTACAGGTTGCTGTCGCGGATCGCGCTTCCGCTTGATTTGTTTCCGCGAGCAAGATCAAC 1307  
Db 1264 GTACCAGATTGCGTGGTCAAGTACAGTGGCTGCTGCTGCTGAGTGGCTTGGTGGCAAG 1427  
QY TCTAAAGTATGGAGTGTGCAGACGAAGAACTGGCACAGAGCTGCTGGACATGGCGGA 1367  
Db 1324 ACTGAAGTGTGGGATGTGAAGGCCACAGAGCTGCCATGGACCTGCCCGCCACGCGGA 1383  
QY TGAGTGTTCGAGTGGCTGGCGCCCGGATGGCTCTAGAGTTGCTTGGTGGCAAGGA 1427  
Db 1384 TGAGTATATGCTGTGAGTGGAGTCCAGATGGCCAGAGAGTGGCAAGTGGTGGGAAGGA 1443  
QY 1428 CAAAGTTATAAGCTATGG 1446  
Db 1444 CAAATGCTCCGGATATGG 1462

## RESULT 13

AX405715 2638 bp DNA linear PAT 14-JUN-2002  
LOCUS Sequence 130 from Patent WO0222660.  
ACCESSION AX405715  
VERSION AX405715.1 GI:21438869  
KEYWORDS human.  
SOURCE Homo sapiens  
ORGANISM Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1  
AUTHORS Tang, Y.T., Liu, C., Zhou, P., Asundi, V., Zhang, J., Zhao, Q.A., Ren, F.,  
Xue, A.J., Yang, Y., Wehrman, T. and Drmanac, R.T.  
TITLE Novel nucleic acids and polypeptides  
JOURNAL Patent: WO 0222660-A 130 21-MAR-2002;  
HYSEQ, INC. (US)

FEATURES  
source

Location/Qualifiers  
1..2638  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
82..1539  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAD34822.1"  
/db\_xref="GI:21438870"  
/translation="MAAADAADAVRDVORLLVQDEGGOLIGSPDVPDVLTPRL  
OLVNCALLAEDPLPLAFVHDIAEIVSLGKTLISQAVETKVLIDLLIQALFVRA  
VTRCTSSLGHESEAVISWAPSPGKVLASGSDTTRVWDLSLEIHPFTCKGRHWL  
STSWSDGKLLASCKNGQILLMDPTSGKOVGRTLAGSKWITGLSWEPLHANPECRY  
VASSKDGSRVLDWTAGRCERILITGHTOSVTLRWGGDGLLYSASQDRTIKYVRAHD  
GLVLCRTQGHGHWNMTALSTYALRTGAFPAEASVNPQDLOGSLQELKERALSRYN  
LVRGGPERLVSGSDDEFTFLWSPADKKPLTRMTGHOALLINQVLFSPDSRIVASAF  
DKSIKLWDGRTGKYLASLRGHVAAYQIAWSADSLRLVSGSSDSTLKVWDVKAQKLA  
MLPGHADEVYAVDWSPDGQVRVSGGDKCLRIWR"

## CDS

54 GATACAGGCGCGCTCGCTTTACACGGCGGAGAGCCGCCCGCAATCGACCTTGCCTGGC 113  
132 GCTAGTGCAGATTCCTCAGGATGAGCGGGCAGCTGTGGGTTCCTCCGACGTGCCCCGT 191  
114 AGGAATCACTACCCAGCAATTTGGAGCTGATTGCAACGGCTGCTGTAACCAAGAGGAGC 173  
192 GGACATCAACCCGGACAGGCTGCAGCTCGTGTGCAACGGCTACTGGCCAGGAGGATCC 251  
174 CACTCCATATTGTTTTCGTCGGCGGAGATGAGATCAAGAAGACCTTGGAGGACAGTT 233  
252 CTTGCCCATGGCTTCTTTTCCAGATGCTGAGATGCTGCTCCTCCTCCTCCTCCTCCTCCT 311  
234 GGACTTGGCGTCACTGGACACCCGAAACGCTGATGATATTTGTTATACGCCACACAGGCG 293  
312 GGAGTCCCAAGGAGTGGAGACAGAGAAAGTCTTAGACATCATCATACAGCCACAGGCTAT 371  
294 TTTCAAAAGTCCGCCAGTGCACAGATGCACAGTTCATGCGCGGAGACACGCCAGGCTGT 353  
372 CTTACAGTCCGGGCTGTGACTCGCTGCACCACTCCTTGGAGGCTCACAGTGAAGGAGT 431  
354 GGTTCGCTGAATTTACGCCCGGATGCTCATCTGCGCAAGTGGCAAGTGGCGACACAC 413  
432 CATTTCTGTGGCTTTCAGCCCTACGGGAAAGTACCTGGCCAGTGGCTCTGGAGACAC 491  
414 AGTGCATTTGTGGGATCTTAACACAGACAGACCGCACTTCACCTGCACAGCTCATAGCA 473  
492 CGTGCCTTCTGGGATCTCAGCAGACAGACACACATTTTCACATGCAAGGACACAGACA 551  
474 GTGGGTTCGTGCTGCTATCCTGGGCTCGGATGGCAACCGTTGGCCAGCGTTGCAAGC 533  
552 CTGGGTCTTAGTATATCTGGTCTCAGATGGCAAGAGTGGCCCTCAGGCTGCAAGAA 611  
534 GGCTCTATATCATCTCTGGACCCGGAGAGCGGTGACGAGAAGGAGGCGACCTTGTAGTG 593  
612 TGCCAGATTTCTCTCTGGGACCCAAAGCAGCAGGAGAGTGGGAGGAGGAGGAGGAG 671  
594 GCACAGAAACACATCAACTCGCTCGCTGGGACCGTATCATCGCATCGGAGTGGAG 653  
672 CCACAGAGTGGATCAGAGCTGAGCTGGGAGCCCTCCATCGAACCTGAGTGGCGG 731  
654 GAACTTGTTCGCGAGTGGAGACGGGAGTGGCGGATTTGGGACCTAAATTTGGGCA 713  
732 CTATGTGGCCAGAGCTCCAAAGATGGCAGTGGGATGCTGGGACACAACTCAGCGCG 791  
714 GTGCTTATGAACATTCGCGACACAAATGCTGTGACAGCTGAGATGGGTGGAGC 773  
792 CTGTGAGCGCATCTCAGCGGACACCCAGTCCGTGCTGCTCGGTGGGAGGGA 851  
774 GGGCTTATTTATACATCTCCAAAGATCGACAGTGAAGATGTGGGAGCAGCTGATGG 833  
852 CGGGCTTCTTACTCTGCTCCAGGACCCGACCATCAAGTCTGAGAGCTCATGACGG 911  
834 AATCTTGTCCGGACGTTCTTGTGCCAAGCTCAGTGGTAAACAACTTGCCTGAGAC 893  
912 TGTGCTGTCCGGACTCTGCAAGGCCACGCCACTGGGTGAACACCATGSCCTCAGCAC 971  
894 CGATTACGCTCTCGGACTGCTTCCATCCATCCGTGAAGATCGCTCGAAGAGCC---A 950  
972 TGACTATGCTCTGCGACTGGGCTTTGAACCTCTGAGCGCTCAGTTAATCCCAAGA 1031  
951 CCTCAGTTTGAAGCAATTCAGGAATTCGCCATCTGCTTGAAGCGCTTACACAGCGCTGTG 1010  
1032 CCTCAAGGATCTTTCAGGAGTTGAAGGAGAGGCTCTCGAGCGGATACACCTCGTGG 1091  
1011 CCCTCAGAGTGGAGTGGTCTGCTTCTGTCGATGACAAACCTCTATCTGTGG--- 1068  
1092 GGGCCAGGCTCCAGAGAGGCTGGTGTCTGGCTCCGACGCTTACCTTATCTCTGTGGT 1151  
1069 -CGGAACAACAGAACAAAGTGGCTTGGAGCGCATGACAGGCGACAGAACTGGTCAACGA 1127  
1152 CCCAGCAGAGGACAAAAGGCTCTCCTCCTGATGACAGGACACCAAGCTCTCATCAACCA 1211  
1128 TGTGAATATTTCGCGGATGTAAAGCTAATTTGGCTGCTGCTTTCATTTGACAGTCACTGG 1187

BASE COUNT 566 a 767 c 764 g 541 t  
ORIGIN

Query Match 30.7%; Score 477.4; DB 6; Length 2638;  
Best Local Similarity 59.8%; Pred. No. 1.3e-108;  
Matches 837; Conservative 0; Mismatches 556; Indels 6; Gaps 2;



Db 1212 GGTGCTTCTCTCTGACTCCCGCATCGTGGCTAGTGCCTCTTTCACAAAGTCCATCAA 1271  
Qy 1188 TCTGTGGCGAGCAGCATGGTCACTACATGCGCACCCTTCGGGGTTCATGTCAGGCTGT 1247  
Db 1272 GCTGTGGGATGGCAGGACGGCAAGTACCTGGCTCCCTACGCGCCACGTGGCTGCCGT 1331  
Qy 1248 TTACAGGTTGCTGTCGCGGACCTCCCGCTGATGTTGCTTCCCGGAGCAAGACTCAAC 1307  
Db 1332 GTACAGATTCGGGTGCTGAGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1391  
Qy 1308 TCTAAAGTATGAGTGTGAGAGCAAGAACTGGCAGGAGCTGCTGGACATCGGA 1367  
Db 1392 ACTGAAGTGTGGATGTGAAGCCAGAGCTGGCCATGGAACCTGCCCGCCACCGCA 1451  
Qy 1368 TGAGGTGTCGGAGTGGAGTGGCGCCGATGGCTCTAGATTTGATTTTCTGCTGGCAAGA 1427  
Db 1452 TGAGGTATATGCTGTGACTGGAGTCCAGATGGCAGAGAGTGGCAAGTGGTGGGAAGA 1511  
Qy 1428 CAAAGTTATAAGCTATGG 1446  
Db 1512 CAAATCGCTCGGATATGG 1530

RESULT 14  
BC018399  
LOCUS  
DEFINITION  
Mus musculus, hypothetical protein similar to beta-transducin  
family, clone MGC:25690 IMAGE:3491925, mRNA, complete cds.  
ACCESSION  
BC018399  
VERSION  
BC018399.1 GI:17390942  
KEYWORDS  
MGC.  
SOURCE  
house mouse.  
ORGANISM  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1. (bases 1 to 1784)  
Strausberg, R.  
Direct Submission  
Submitted (03-DEC-2001) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgabps-remail.nih.gov](mailto:cgabps-remail.nih.gov)  
Tissue Procurement: Gilbert Smith, Ph.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILLNL)  
DNA Sequencing by: Baylor College of Medicine Human Genome  
Sequencing Center  
Center code: BCM-HGSC  
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>  
Contact: [amg@bcm.tmc.edu](mailto:amg@bcm.tmc.edu)  
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,  
Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,  
Richards, S., Gibbs, R.A.  
Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/ILLNL at: <http://image.llnl.gov>  
Series: IRAK Plate: 30 Row: b Column: 14.  
Location/Qualifiers  
1. .1784  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/map="CZECH II"  
/clone="MGC:25690 IMAGE:3491925"  
/tissue\_type="Mammary tumor metastasized to lung. Tumor  
arose spontaneously from a senescent normal mammary  
(clonal) outgrowth infected with the virus MTV."  
/clone\_lib="NCI\_CGAP\_Lu29"  
/lab\_host="DH10B"  
/note="Vector: pCMV-SPORT6"

FEATURES  
source

CDS  
31..1488  
/codon\_start=1  
/product="hypothetical protein similar to beta-transducin  
family"  
/protein\_id="AAH18399.1"  
/db\_xref="GI:17390943"  
/db\_xref="LocusID:54475"  
/translation="MAAAVFEAAAGDVORLLVQFODEGQOLLGSPFDPVVDITPPDKL  
QLVNCALLAQEPLPLAIVYDAIEVSSLGKLEQSVETEKIVDIYQPOAVFRVA  
VTCTSSLHGESEAVISVAFSTGKILASGSDTTVRFWDLSTPHTCKGHRHWL  
STWSPDGKKLASGCKNGQILLMDPSTGLQVGRILTGHSKWLITGLSWPLHMECHRY  
VASSKDGSRVYMDTARGCEBRILITGHTSVTCCLRWGGDLLISASQDITLKVRAHD  
GVLCRTLOGHGHVNTMALSTYALRTGAPEPAETVNAQDLQGLSKELKRRASRYN  
LVFGQCPERLVSGSDFTFLFWSPAEDKPLRMTQHQLINOVLFSPDSIVASAF  
DKSIKLWDRTCKYKLASLRHVAAYQIAWSADSRLLVSGSDSTLTKVMDVKAQKLAT  
DLPGHADVEYAVDWSPDGQORVASGGKDKLRIWR"

BASE COUNT 397 a 521 c 539 g 327 t

ORIGIN  
Query Match 30.6%; Score 476.6; DB 10; Length 1784;  
Best Local Similarity 59.2%; Pred. No. 1.9e-108;  
Matches 850; Conservative 0; Mismatches 579; Indels 6; Gaps 2;  
Qy 18 GGAGACGACACGAGCAGGACGACGACATACAGATACAGCGCGCGCTGTTTACAC 77  
Db 45 GGTGAAGAGCGCGCGAGCGACGTGCGAGCGCTGCTCGTACATGTTCCAGGATGAGG 104  
Qy 78 GGGCGAGAGCG 137  
Db 105 CGGTCAAGCTGCTGGGCTCCCGTTCGAGCTGCGCGCGCGCGCGCGCGCGCGCG 164  
Qy 138 ACTGATTTGCAACGCGCTGCTGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 197  
Db 165 GCTGTGTGCAACGCGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 224  
Qy 198 CGAGATGATGATCAAG 257  
Db 225 CGATGCTGAGATGTTGTCCTCCCTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 284  
Qy 258 AAGCGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 317  
Db 285 GAAATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 344  
Qy 318 ATGACAGAGTTTCATGCGCGGACACGCGAGCGTGTGTTTTCGCTGAAATTCAGCG 377  
Db 345 GTGCACACGCTCATTTGGAGGCTCAGACGAGCTGCTATTTCTGTGCGCTTCAGCG 404  
Qy 378 TGGTGCTCATTCGCCAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAG 437  
Db 405 AGGAAAGTACCTGGCGCAGTGGTTCGGAGACACACACTGTGCGCTTCTGGGATCTCAG 464  
Qy 438 AGAGACACGCGACTTCACCTGACAGGTGATGAGTGGAGTGGAGTGGAGTGGAGTGG 497  
Db 465 TGAGACACGCGACTTCACATGCAAGAGACACCGCGCGCGCGCGCGCGCGCGCGCG 524  
Qy 498 TCCGATGGCAAGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 557  
Db 525 CCCAGATGGCAAGAACTGGCGCTCAGGCTGCAAGAAAGTGGCGAGAGTTCGCTGTTGG 584  
Qy 558 GGAGACGCGGTTCAGCAG 617  
Db 585 AAGCAGCGGCTGCGAGTGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 644  
Qy 618 CGCCTGGGAACCGTATATCGCATCCGGAGTGGAGGAGGAGGAGGAGGAGGAGGAGGAG 677  
Db 645 GAGCTGGGAGCGCCCTTCACATGAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 704  
Qy 678 CGGGAGCTCGCGGATTTGGGACGTAATAATTTGGCGCGCGCGCGCGCGCGCGCGCG 737  
Db 705 CGGCGAGTACAGCTCTGGGATACAAACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 764  
Qy 738 CACAAATGCTGTGACAGCAGTGAAGTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 797

Db 765 CAGCAGTCACTCAGTCCGCTCCGATGGGAGGAGATGGCTTCTATATTTCTTCCCA 824  
QY 798 AGATCGCAGTGAAGATGTGGGAGCAGCTGATGAATCTTGTGGGAGCTTCTCTGG 857  
Db 825 GGACCGACCACTCAAGCTGTGGAGGCCCATGACGGTGTGCTGTGCTACTTCAAGG 884  
QY 858 CCAAGCTCACTGGGTAAACAACATTCGCTGAGCACCAGATTACGTCCTCGCAGCTGGTCC 917  
Db 885 TCATGCCCACTGGTGAACACCATGCGCTTAGCACAGACATATGCCCTGGCAGAGGGC 944  
QY 918 ATTCCATCGGTGAAGATCGCTCCAAGAGC--ACCTCAGTTGACGACTGAGGAAT 974  
Db 945 CTTTGAGCTCTGAAGCCACAGTGAATGCCCAAGACCTGCGAGGGTCTTGAAGAGCT 1004  
QY 975 GCAGGAATCTCCCTTGAAGCGCTACAGCGCTGTGCGCTGTGACGAGGTGGAGTCTGGT 1034  
Db 1005 GAAGGAGAGGCACTCGAGCGCTTACAACTCTGTGCGGGGCCAAGGCCAGAGAGCTGT 1064  
QY 1035 TTCTGTGTCGATGACAAACACCTCTATCTGTGG---CGGAACAACAGCAAGTGGT 1091  
Db 1065 CTCTGGCTCAGACGACTTCACTTATCTGTGTGTCCTGAGCAGGAGCAAGACCCCT 1124  
QY 1092 TGAGCGCATGACAGCGCAGCAGACGTTGCTCAACGATGTGAATATTCGCCGGATGTAA 1151  
Db 1125 CCCCCGATGACAGGACACAGGCCCTCATCAACAGGTCTCTCTCTCTGACTGCG 1184  
QY 1152 GCTAATGCGCTGCTTCAATTTGACAAGTCACTGCTGTGGCGAGCAGCGATGTCA 1211  
Db 1185 CATCTGGCCAGTCTCTCTGACAAAGTCTATCAAGCTGTGGATGTGGCGACGGCA 1244  
QY 1212 GPACATGCCACCTTCGGGTCATGTGCAGCTGTTCACGGTTCCTGTGTCGGGA 1271  
Db 1245 GTACCTAGTCTCTGCTGGCCAGCTGGCAGCGGTGTACCAATGTGCTGTGACTGGA 1304  
QY 1272 CTCGCGCTGTGTTTCCGCGAGCAAGACTCAACTCTAAAAGTATGAGTGTGCAGC 1331  
Db 1305 CAGCGCATATTAGTCAGCGGAGCAGTGCAGCAGCTGAAGGTGTGGATGTGAAGC 1364  
QY 1332 GAAGAAGTGCACAGGAGCTGCTGGACATCGGATGAGTCTTGGAGTGTGAGTGGC 1391  
Db 1365 CCAGAACTGCCACCGAGCTGCTGGCCATGCGATGCGATGAGTATGCTGTGACTGG 1424  
QY 1392 GCCGATGCTCTAGAGTGTCTCTGTGGCAAGGACAAAGTTATTAAGCTATGG 1446  
Db 1425 TCCAGATGGCCAGAGATAGCAAGCGCGGGAAGACAAGTGTCTCGGATATGG 1479

## RESULT 15

AXI35807  
LOCUS AXI35807 1928 bp DNA linear PAT 29-MAY-2001  
DEFINITION Sequence 1 from Patent WO0132614.  
ACCESSION AXI35807  
VERSION AXI35807.1 GI:14272042  
KEYWORDS human.  
SOURCE Homo sapiens

ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE  
1 (bases 1 to 1928)  
AUTHORS Utku,N.

TITLE Novel genes tzap7/a, tzap7/b and tzap7 involved in t cell

activation and uses thereof

JOURNAL Patent: WO 0132614-A 1 10-MAY-2001;

Utku, Nalan (DE)

FEATURES  
source Location/Qualifiers

1..1928

/organism="Homo sapiens"

/db\_xref="taxon:9606"

92..1546

/note="unnamed protein product"

/codon\_start=1

/protein\_id="CAC39574.1"

/db\_xref="GI:14272043"

/translation="MWIADAVARDVQRLLVQFQDEGGQLLGSFFDVPVDIIPDRIQ  
LVNALLAQEDPLSLAFFVHDAEIVSLGKTLSEAVETEKVLDIILPTQKVFVRAV  
TRCTSLSEKHTAEVIVAFPTGKIVLASGGDTTVRFWDLSTETPHTSKGHTFWLVS  
IAMSPPDKKILASGCKNSQIFIWDPSTCKQIKGPKLTHGSKWLLHNPENRYL  
ASASSGRVDRITWDTAGRCERILITGHTSVTCRLRWGGDGLLYSASODRTIKVRAHDG  
VLCRTLOGHGHVNTMALSTDYALRTGAPEAPASVNPQDLOGSLQELKERALSRYNL  
VRGQPERLVSGSDDFTLFLWSPAEDKKPLRTWGHQALINOLFLSPDSRIVASFD  
KSLKLDERTCKYLASLRGHVAAYQIANSADSRLLVSGSSDSTLKVWDVKAOKLAMD  
LPGHAEVYAVDPSDQORVASGGKDLRIWR"

BASE COUNT 432 a 540 c 550 g 406 t

ORIGIN

Query Match 30.0%; Score 466.2; DB 6; Length 1928;

Best Local Similarity 59.3%; Pred. No. 7.7e-106;

Matches 830; Conservative 0; Mismatches 563; Indels 6; Gaps 2;

QY 54 GATACAGGCGCCCTCGTTTACACGGCGGAGGAAGCCGCCGCAATCGACCTGCCGC 113  
Db 139 GCTAGTGCAGTTCAGGATGAGGCGGGCAGCTGTGGGTTCCCCGTTCCGACGTG 198  
QY 114 AGGAATCACTACCCAGCAATTTGGGACTGATTTGCAACGGCTGTCTGAAAAACGAGGA 173  
Db 199 GGACATCAACCCGGACAGCGCTGCAGCTCGTTTGCACCGGCTACTTGGCCAGGAGTCC 258  
QY 174 CACTCCATATTTGTTTTTCGTGGCGGAGATGAGATCAAGAGAGCCCTGGAGGACAGTT 233  
Db 259 CCTGTCACTGGCTTTCTTTGTCACGATGCTGAGATCTTCTCTCACTCGGGAAGAGTT 318  
QY 234 GGACTTGGCGTCAGTGGACACCGGAAACGTGATCGATATTGTGTATACGCAAGCGGT 293  
Db 319 GGAGTCCCAGGCGAGGAGAGAGAGAGTCTTAGACATCATTTTACCACACAAAGCTGT 378  
QY 294 TTTCAAAGTGCGCCCGAGTGCACAGATGCAGAGTTCATTCGCGGAGACACGCGAGGCTGT 353  
Db 379 GTTCAAGTTCTGCTGTGTACACAGATGTACCAGCTCATTTGGAGGAGACACACCGAGCTGT 438  
QY 354 GGTTCGCTGATTTTCACGCCCGGATGGTCTCATCTCGCGAGTGGAGTGGGACACAC 413  
Db 439 TATTTCACTAGCTTTCAGCCCACTTGGAAAGTATTTGGCAAGTGTGTTCTGGGACACTAC 498  
QY 414 AGTGGATTTGGGATCTTAACACAGAGACCGGACTTCACTTCACCTGCACAGTCTAAGCA 473  
Db 499 AGTCCGCTTTTGGATCTCAGACAGAAATCCACATTTTACATCTAAAGGATACACA 558  
QY 474 GTGGGTCTGTGCTGTATCTCGGCTCGGATGCAAGAGCTTTGGCCAGCGGTGCAAGC 533  
Db 559 CTGGGTCTCAGTATGCTTGTCTCCAGATGGCAAAAACCTTGCCTCAGGATGTAAAA 618  
QY 534 GGGCTCTATATATCTTGGACCGGAGAGCGGTGAGCAAGGGGGGACCTTGTAGTGG 593  
Db 619 TAGTCAGATCTTCTATTTGGGACCCCAAGCAGGGAAGCAGATTTGGCAAAACCATTAACAG 678  
QY 594 GCACAAGAAACACATCAACTCGCTCGCTCGGAAACCGTATCATCGCGATCCGAGTGCAG 653  
Db 679 GCACCTCAAAGTGGATTACATGGCTGTGTTGGGAACCTCTCCACCTGAACCCAGAGCG 738  
QY 654 GAAACTGCTTCCCGCAGTGCAGACCGGAGTCCCGGATTTGGGACCTAAATTTGGGCA 713  
Db 739 ATACCTACCGAGTGCCTCCAGCGCGCGCTCGCGGATCTGGGACACAACTCGAGCGCG 798  
QY 714 GTGCTTATGAACATTTCCGCGACACAAATCTGTGACAGAGTGGAGTGGGTGGAGC 773  
Db 799 CTGTGAGCGCATCTCTACCGGCGACACCCAGTCCGCTACCTGTCTCGGTGGGAGGGGA 858  
QY 774 GGGCTCTATTTATATCTCTCAAGATCGCAGTGGAGATGTGGGAGCAGCTGATGG 833  
Db 859 CGGGCTTCTTCTCTCTCCCGAGGCGCGACCATCAAGTCTGGAGAGTCTATGACGG 918  
QY 834 AATCTGTGCGGAGCTTCTCTGGCCAGCTCAGTGGTAAACAACTATGCGTGGAGCAG 893  
Db 919 TGTGCTGTGCGGAGCTTCTGCAAGGCCACCGCCACTGGGTGAACACCATGSCCTCAGCAG 978  
QY 894 CGATTACGCTCTGCGCAGCTGGTCTCATTCATCCGCTGGAAGGATGCGCTCCAGAGCC ---A 950

Db	979	TCACATATCCCTGGCCACTGGGCGCTTTGAACCTGCTGAGGCGCTCAGTTAATCCCCAAGA	1038
Qy	951	CCTCAGTTTGGAGCACTGAGGAATTCGAGGAATCTGCCTTGAAGCGCTACCAAGCGCGTGTG	1010
Db	1039	CCTCAAGATCCCTGAGGAGTTGAAGGAGAGGCTCTGAGCGGATACAACCTCGTGG	1098
Qy	1011	CCCTGACAGAGTGGAGTCGTGTTCCCTGTTGCGATGACAAACCTCTATCTGTGG--	1068
Db	1099	GGCCAGAGGTCACAGAGGCTGTGCTGGCTCCGACACTTCACCTTATCTCTGTGGTC	1158
Qy	1069	-CGGAACAACAGAACAAAGTGCCTTGAGCGCATGACAGGCGACCAAGACGTCGTCAACGA	1127
Db	1159	CCCAGCAGAGGACAAAAGCCTCTCACTCGGATGACAGACCAAGCTCTCATCAACCA	1218
Qy	1128	TGTGAATATTCGCGGGATGTAAGCTAATTGCGTCTGCTTCATTTCACAAGTCAGTGG	1187
Db	1219	GGTGCTCTCTCTGACTCCCGCATCGTGGCTAGTGCCTCTTGCACAGTCCATCAA	1278
Qy	1188	TCTGTGGGAGCCAGCGATGGTCAGTACATGCCACCTTCCGGGGTCATGTGAGGCTGT	1247
Db	1279	GCTGTGGGATGGCAGGACGGGCAAGTACCTGGCTTCCCTACGCGGCCACGTGGCTGCCGT	1338
Qy	1248	TTACAGGTTGCTGGTCCCGGACTCCCGCTTGATTGTTTCCGCGCAGCAAGACTCAAC	1307
Db	1339	GTACCAGATTGGTGGTCAGCTGACGTGGCTCGGCTCTGTCAGCGGCGAGCTGACAGCAC	1398
Qy	1308	TCTAAAGTATGGAGTGTGACAGAGAACTGGCAGAGAGCTGCTGGACATCGCGGA	1367
Db	1399	ACTGAAGTGTGGGATGTGAAGGCCAGAGCTGGCCATGGACCTGGCCGCCACGCGGA	1458
Qy	1368	TGAGGTGTTGGAGTGGACTGGGCGCCGATGGCTCTAGAGTTGCTCTGTGTGCAAGGA	1427
Db	1459	TGAGGTATATGCTGTTGACTCCAGATGCCAGAGAGTGGCAAGTGGTGGGAAGGA	1518
Qy	1428	CAAAAGTTATAAGCTATGG	1446
Db	1519	CAAAATGCCTCCGGATATGG	1537

Search completed: June 2, 2003, 04:13:27  
Job time : 3999 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: June 2, 2003, 01:13:14 ; Search time 378 Seconds  
(without alignments)  
9264.173 Million cell updates/sec

Title: US-09-830-980-2  
Perfect score: 1555  
Sequence: 1 aattcccaaaatgcagga.....aaaaaaaaaaaaaaaaaaaa 1555

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_101002:\*

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:\*

2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:\*

3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:\*

4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:\*

5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:\*

6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:\*

7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:\*

8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:\*

9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:\*

10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:\*

11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:\*

12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:\*

13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:\*

14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:\*

15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:\*

16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:\*

17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:\*

18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:\*

19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:\*

20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:\*

21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:\*

22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:\*

23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:\*

24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	1555	100.0	1555	21 AAA27739
2	1487.8	95.7	1992	23 ABL05339
3	1285.8	82.7	4164	23 ABL05338
4	601.2	38.7	6223	23 ABL05232
5	588.2	37.8	4141	23 ABL05214
6	478	30.7	1859	22 AAF85075
7	477.4	30.7	1853	22 AAF85075
8	477.4	30.7	1854	22 AAF85075
9	477.4	30.7	2638	24 ABL05339

10	466.2	30.0	1928	22 AAF85073
11	333.4	21.4	1726	21 AAC48725
12	331.8	21.3	1643	21 AAC39906
13	249.4	16.0	1170	22 AAF85074
14	176	11.3	399	21 AAF85071
15	164	10.5	547	22 AAH05727
16	155.8	10.0	1722	20 AAX30340
17	153.6	9.9	573	22 AAF98017
18	123.8	8.0	540	21 AAF07777
19	120.4	7.7	378	22 AAF67530
20	80.4	5.2	316	21 AAF11429
21	79.4	5.1	2984	22 AAK94739
22	79.4	5.1	3025	22 ABA06363
23	78.4	5.0	300	20 AAZ14360
24	78.4	5.0	764	22 AAK91829
25	78.4	5.0	764	22 AAK93229
26	78.4	5.0	1950	22 AAK52860
27	69	4.4	1925	22 AAK51876
28	68.4	4.4	279	23 AAV55135
29	68.4	4.4	2977	23 ABO72680
30	68.2	4.4	1464	24 ABL59758
31	68.2	4.4	1751	22 AAF58329
32	68.2	4.4	2064	21 AAF23275
33	64	4.1	1591	23 ABL23519
34	62.6	4.0	462	23 ABL57002
35	62.2	4.0	559	23 ABL54541
36	62	4.0	1810	21 AAC77522
37	61.8	4.0	2359	15 AAF070726
38	61.4	3.9	821	22 AAK92334
39	61.4	3.9	821	22 AAK93787
40	61.4	3.9	1890	22 AAK94610
41	60.2	3.9	2359	17 AAT42212
42	60.2	3.9	2359	18 ABL79603
43	60.2	3.9	2826	23 ABL13261
44	60.2	3.9	3965	23 ABL11658
45	60.2	3.9	4826	23 ABL13260

ALIGNMENTS

RESULT 1  
AAA27739  
ID AAA27739 standard; cDNA; 1555 BP.  
AC AAA27739;  
XX 29-AUG-2000 (first entry)  
DT XX Drosophila Notchless (Nle) cDNA.  
DE XX Notchless; Nle gene; Notch; signalling; neurodegenerative disease;  
KW cancer; diagnosis; cytostatic; neuroprotective; gene therapy; ss.  
XX OS Drosophila melanogaster.  
XX FH Key  
XX CDS Location/Qualifiers  
FT 13..1455  
FT /\*tag=a  
FT /transl\_except= (pos:859..861, aa:His)

WO200026364-A1.  
PD 11-MAY-2000.  
XX 03-NOV-1999; 99WO-IB01891.  
XX 03-NOV-1998; 98GB-0024045.  
XX (EUMO-) EURO MOLECULAR BIOLOGY LAB.  
XX Cohen S, Bouwmeester A, Royet J;  
XX

DR WPI: 2000-365613/31.  
 XX P-PSDB: AAV79678.  
 PT Novel Notchless protein and nucleic acids encoding them useful for  
 PT treating and preventing cancer and neurodegenerative diseases  
 XX  
 PS Claim 19; Page 44-45; 52pp; English.  
 XX  
 CC The present sequence is that of cDNA encoding Notchless, a novel  
 CC protein of Drosophila. The cDNA was isolated from an eye disc cDNA  
 CC library by screening with a genomic DNA fragment that flanks the P  
 CC element insertion site in Drosophila 1(2)AAK13714. Notchless was  
 CC identified in a screen for dominant modifiers of a Notch mutant  
 CC phenotype in the Drosophila wing. The mutant dominantly suppressed  
 CC the wing notching phenotype of notchoid mutations, and the Notchless  
 CC protein was shown to bind to the cytoplasmic domain of Notch.  
 CC Notchless modified Notch signalling activity in a variety of  
 CC embryos. The Notchless protein, and nucleic acids encoding it, can  
 CC be used in methods for the diagnosis and therapy of certain diseases,  
 CC particularly cancer and neurodegenerative diseases (claimed). A  
 CC probe capable of screening for the Notchless gene (Nle), a cloning  
 CC or expression vector comprising Notchless DNA, cDNA or RNA, a host  
 CC cell, a transgenic animal, and a method for production of Notchless  
 CC protein in the host cell are also claimed.  
 XX  
 SQ Sequence 1555 BP; 418 A; 373 C; 448 G; 316 T; 0 other;

Query Match 100.0%; Score 1555; DB 21; Length 1555;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1555; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

601 AAACACATCACTGCTGCGCTGGGAACCGTATCATCGCATCCGAGTCGAGAACTT 660  
 661 GCCTCCCGCAGTGGAGACGGGACTCCCGATTTGGAGTAAATTTGGAGTGGCCAGTGGCTT 720  
 661 GCCTCCCGCAGTGGAGACGGGACTCCCGATTTGGAGTAAATTTGGAGTGGCCAGTGGCTT 720  
 721 ATGAACATTTCCCGGACACACAAATGCTGTGACAGCAGTGGATGGGGTGGAGCGGGCTT 780  
 721 ATGAACATTTCCCGGACACACAAATGCTGTGACAGCAGTGGATGGGGTGGAGCGGGCTT 780  
 781 ATTATACATCTCTCAAAAGATCGCAGAGTGAAGATGTCGCGAGCAGTGTATGGAACTTGG 840  
 781 ATTATACATCTCTCAAAAGATCGCAGAGTGAAGATGTCGCGAGCAGTGTATGGAACTTGG 840  
 841 TSCCGGACGCTTCTCTGGCCAAAGCTCACTGGGTAAACAACTTGGCGCTGAGCACCAGTAC 900  
 841 TSCCGGACGCTTCTCTGGCCAAAGCTCACTGGGTAAACAACTTGGCGCTGAGCACCAGTAC 900  
 901 GTCTCGCAGCTGGTCCATTCATCCCGGTGAAGGATGCGTCCAGAGCAGCCTCAGTTTG 960  
 901 GTCTCGCAGCTGGTCCATTCATCCCGGTGAAGGATGCGTCCAGAGCAGCCTCAGTTTG 960  
 961 AGCAGTGAAGGATTCAGGATCTGCTTGAAGCGGTACAGCGGCTGTCGCTGACGAG 1020  
 961 AGCAGTGAAGGATTCAGGATCTGCTTGAAGCGGTACAGCGGCTGTCGCTGACGAG 1020  
 1021 GTGAGTGCCTGCTTCTCTGGATGACAAACACCTCTATCTGTGCGGAAACAAACAG 1080  
 1021 GTGAGTGCCTGCTTCTCTGGATGACAAACACCTCTATCTGTGCGGAAACAAACAG 1080  
 1081 AACAGTGGTGTGAGGCGATGACAGGCGCAGGAGTGTGCAAGTGTGAATATTCG 1140  
 1081 AACAGTGGTGTGAGGCGATGACAGGCGCAGGAGTGTGCAAGTGTGAATATTCG 1140  
 1141 CGGATGTAAAGCTAAATGCGTCTGCTTCAATTTGACAAAGTCACTCACTTAAAGTATGG 1200  
 1141 CGGATGTAAAGCTAAATGCGTCTGCTTCAATTTGACAAAGTCACTCACTTAAAGTATGG 1200  
 1201 AGCGATGTGATGAGTGGCGACCTTCCGGGGTCACTGTGCGAGGCTGTTTACAGGTTGG 1260  
 1201 AGCGATGTGATGAGTGGCGACCTTCCGGGGTCACTGTGCGAGGCTGTTTACAGGTTGG 1260  
 1261 TGGTCCGCGGACTCCCGCTTGAATGTTTCCGGCGACCAAGACTCACTTAAAGTATGG 1320  
 1261 TGGTCCGCGGACTCCCGCTTGAATGTTTCCGGCGACCAAGACTCACTTAAAGTATGG 1320  
 1321 AGTGTGACAGCAAGAACTGGCAGAGGAGTGCCTGGACATGCGGATGAGGTGTCGGA 1380  
 1321 AGTGTGACAGCAAGAACTGGCAGAGGAGTGCCTGGACATGCGGATGAGGTGTCGGA 1380  
 1381 GTGAGTGGGCGGCGGCTGCTAGAGTGGCTTGGCTGCTGCTGGCAGCAAGAACTTAAAG 1440  
 1381 GTGAGTGGGCGGCGGCTGCTAGAGTGGCTTGGCTGCTGCTGGCAGCAAGAACTTAAAG 1440  
 1441 CTATGGGCTTATTAAACAAATCAATTAACCTGTACAGCGTAAAGAACTTAAAG 1500  
 1441 CTATGGGCTTATTAAACAAATCAATTAACCTGTACAGCGTAAAGAACTTAAAG 1500  
 1501 TAAACGCTCTGAGTAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC 1555  
 1501 TAAACGCTCTGAGTAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC 1555

## RESULT 2

ABL05339  
 ID ABL05339 standard; cDNA; 1992 BP.

XX ABL05339;

XX AC  
 XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 10499.

XX	Drosophila; developmental biology; cell signalling; insecticide;	QY	424	TGGGATCTTAAACACAGACACCGCAGCTTCACCTGCACAGGTTCATACGAGTGGTCTCG	483
KW	pharmaceutical; gene; ss.	DB	439	TGGGATCTTAAACACAGACACCGCAGCTTCACCTGCACAGGTTCATACGAGTGGTCTCG	498
XX	Drosophila melanogaster.	QY	484	TGCGTATCCTGGGCTCCCGATGCGCAACGGTTGGCCAGCGGTTCGCAAGCGGGCTCTATA	543
XX	WO200171042-A2.	DB	499	TGCGTATCCTGGGCTCCCGATGCGCAACGGTTGGCCAGCGGTTCGCAAGCGGGCTCTATA	558
XX	27-SEP-2001.	QY	544	ATCATCTGGGACCCCGAGACGGGTCACGAGAAGGGCGACCTTCAGTGGGCAACAGAA	603
XX	23-MAR-2001; 2001WO-US09231.	DB	559	ATCATCTGGGACCCCGAGACGGGTCACGAGAAGGGCGACCTTCAGTGGGCAACAGAA	618
XX	23-MAR-2000; 2000US-191637P.	QY	604	CACATCAACTGCCTCGCTGGGAACCCGTATCATCGGATCCGAGATGCGAGAACTTGCT	663
PR	11-JUL-2000; 2000US-0614150.	DB	619	CACATCAACTGCCTCGCTGGGAACCCGTATCATCGGATCCGAGATGCGAGAACTTGCT	678
XX	(PEKE ) PE CORP NY.	QY	664	TCGCGCAGTGGAGACGGGACTGCGGATTTGGAGCTAAATTTGGGCGAGTGCCTTATG	723
XX	Venter JC, Adams M, Li PWD, Myers EW;	DB	679	TCGCGCAGTGGAGACGGGACTGCGGATTTGGGACGTGAATTTGGGCGAGTGCCTTATG	738
XX	WPI; 2001-656860/75.	QY	724	AACATTCGCGGACACAAATGCTGTGACAGCAGTGAATTTGGGCGAGTGCCTTATG	783
DR	P-PSDB; ABB61236.	DB	739	AACATTCGCGGACACAAATGCTGTGACAGCAGTGAATTTGGGCGAGTGCCTTATG	798
XX	New isolated nucleic acid detection reagent for detecting 1000 or more	QY	784	TATACATCTCTCCAAAGATCGCACAGTGAAGATGTGGCGAGCAGCTGATGGAATCTTGTC	843
XX	genes from Drosophila and for elucidating cell signalling and cell-cell	DB	799	TATACATCTCTCCAAAGATCGCACAGTGAAGATGTGGCGAGCAGCTGATGGAATCTTGTC	858
XX	interactions -	QY	844	CGGAGCTTCTGGGCAACGCTCAGTGGTGAACAACTTGGCTGAGCAGCAGTATACGTC	903
XX	Claim 1; SEQ ID NO 10499; 21pp + Sequence Listing; English.	DB	859	CGGAGCTTCTGGGCAACGCTCAGTGGTGAACAACTTGGCTGAGCAGCAGTATACGTC	918
CC	The invention relates to an isolated nucleic acid detection reagent	QY	904	CTGCGCACTGGTCCATTCATCCGCTGAAGGATGCTGCCAAGAGCGACCTCAGTTGAGC	963
CC	capable of detecting 1000 or more genes from Drosophila. The invention is	DB	919	CTGCGCACTGGTCCATTCATCCGCTGAAGGATGCTGCCAAGAGCGACCTCAGTTGAGC	978
CC	useful in developmental biology and in elucidating cell signalling and	QY	964	ACTCAGGAATTCGAGCAATCTGCTTGAAGCGCTTACAGCGCGTGTGCGCTCAGCAGGTG	1023
CC	cell-cell interactions in higher eukaryotes for the development of	DB	979	ACTCAGGAATTCGAGCAATCTGCTTGAAGCGCTTACAGCGCGTGTGCGCTCAGCAGGTG	1038
CC	insecticides, therapeutics and pharmaceutical drugs. The invention	QY	1024	GAGTGGTGTTCCTGTTGGATGACAAACCTCTATCTGTGGGCGGACCAACAGAAC	1083
CC	discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA	DB	1039	GAGTGGTGTTCCTGTTGGATGACAAACCTCTATCTGTGGGCGGACCAACAGAAC	1098
CC	sequences (ABL01840-ABL16175) and the encoded proteins	QY	1084	AAGTGGCTTCAGCGCATGACAGGCGCACCAAGCTGTGTAACGATGTGAATATTCGCGC	1143
CC	(ABB57737-ABB72072).	DB	1099	AAGTGGCTTCAGCGCATGACAGGCGCACCAAGCTGTGTAACGATGTGAATATTCGCGC	1158
CC	The sequence data for this patent did not form part of the printed	QY	1144	GATGTAAGCTAATTCGCTTCTGCTTCAATTCGCAAGTCAGTGTCTGTGGCGAGCCAGC	1203
CC	specification, but was obtained in electronic format directly from WIPO	DB	1159	GATGTAAGCTAATTCGCTTCTGCTTCAATTCGCAAGTCAGTGTCTGTGGCGAGCCAGC	1218
CC	at ftp.wipo.int/pub/published_pct_sequences.	QY	1204	GATGTCAGTACATGCGCCACCTTCGCGGGTTCATGTGAGGCTGTGTACAGGTTGCTGCG	1263
XX	Sequence 1992 BP; 482 A; 492 C; 585 G; 433 T; 0 other;	DB	1219	GATGTCAGTACATGCGCCACCTTCGCGGGTTCATGTGAGGCTGTGTACAGGTTGCTGCG	1278
XX	Query Match 95.7%; Score 1487.8; DB 23; Length 1992;	QY	1264	TCGCGGAGCTCCCGCTGTGATTTCCGCGAGCAAGAGCTCAACTCTAAAGTATGAGGT	1323
XX	Best Local Similarity 99.3%; Pred. No. 0;	DB	1279	TCGCGGAGCTCCCGCTGTGATTTCCGCGAGCAAGAGCTCAACTCTAAAGTATGAGGT	1338
XX	Matches 1505; Conservative 0; Mismatches 7; Indels 3; Gaps 1;	QY	1324	GTGCAGAGCAAGAACTGCGCACAGGAGTGCCTTGACATGCGGATGAGGTGTTCCGAGTG	1383
QY	7 CAAAAAATCAGGAGACGGACAGGAGCGAGAGCGCCACATACATACAGGCGCC 66	DB	1339	GTGCAGAGCAAGAACTGCGCACAGGAGTGCCTTGACATGCGGATGAGGTGTTCCGAGTG	1398
DB	19 CAGAAAATCAGGAGACGGACAGGAGCGAGAGCGCCACATACATACAGGCGCC 78	QY	1384	GACTGGGCGCCGATGCTCTAGAGTTGCTCTGTTGGCAAGCAAGTATTAAGCTA	1443
QY	67 CTCGTTT---ACACGGCGAGGAAGCGCGCCGCAATCGACCTGCGCGAGGAATCACT 123	DB	1399	GACTGGGCGCCGATGCTCTAGAGTTGCTCTGTTGGCAAGCAAGTATTAAGCTA	1458
DB	79 CTCGTTTCGGACACGGCGAGGAAGCGCGCCGCAATCGACCTGCGCGAGGAATCACT 138	QY	1444	TGGGCTTATTAACAAATCATTAATGTTACAGCGTAAGAAATTAAGTAAGTAA	1503
QY	124 ACCAGCAATTCGGAGTATTTGCAAGCGCTGCTGAAAACGAGGAGCCACTCCATAT 183	DB	1459	TGGGCTTATTAACAAATCATTAATGTTACAGCGTAAGAAATTAAGTAAGTAA	1518
DB	139 ACCAGCAATTCGGTCTGATTTGCAAGCGCTGCTGAAAACGAGGAGCCACTCCATAT 198				
QY	184 TTGTTTTTCGTCGGCGAGGATGAGATCAAGAAGAGCTGGAGAGACAGTTGGACTTGGCG 243				
DB	199 TTGTTTTTCGTCGGCGAGGATGAGATCAAGAAGAGCTGGAGAGACAGTTGGACTTGGCG 258				
QY	244 TCAGTGGACACCGAAACGTGATGATTTGCTATTCAGCCACAGCGGTTTCAAGTG 303				
DB	259 TCAGTGGACACCGAAACGTGATGATTTGCTATTCAGCCACAGCGGTTTCAAGTG 318				
QY	304 CGCCAGTGCAGAGATGCACGAGTTCATCCCGGGACACCGCGAGGCTGTGGTTGCGTG 363				
DB	319 CGCCAGTGCAGAGATGCACGAGTTCATCCCGGGACACCGCGAGGCTGTGGTTGCGTG 378				
QY	364 AATTTCAGCCCGGATGCTCATCTCGCCAGTGAAGTGGCGACACACAGTCCGATG 423				
DB	379 AATTTCAGCCCGGATGCTCATCTCGCCAGTGAAGTGGCGACACACAGTCCGATG 438				







PN WO200171042-A2.  
 XX 27-SEP-2001.  
 XX 23-MAR-2001; 2001WO-US09231.  
 XX 23-MAR-2000; 2000US-191637P.  
 XX 11-JUL-2000; 2000US-0614150.  
 XX (PEKE ) PE CORP NY.  
 XX Venter JC, Adams M, Li PWD, Myers EW;  
 XX WPI; 2001-656860/75.  
 XX P-PSDB; ABB61111.  
 XX New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX Claim 1; SEQ ID NO 10124; 21pp + Sequence Listing; English.  
 XX The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABLI16176-ABLI30511), expressed DNA  
 CC sequences (ABLI01840-ABLI16175) and the encoded proteins  
 CC (ABB57737-ABB72072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX Sequence 4141 BP; 982 A; 1023 C; 1048 G; 1088 T; 0 other;  
 SQ

Query Match 37.88; Score 588.2; DB 23; Length 4141;  
 Best Local Similarity 81.9%; Pred. NO. 1.4e-122;  
 Matches 792; Conservative 0; Mismatches 3; Indels 172; Gaps 3;

QY 724 AACATTGCGGACACACAAATGCTGTACAGCAGTGAAGTGGGCGGCGCTTATT 783  
 DB 1 AACATTGCGGACACACAAATGCTGTACAGCAGTGAAGTGGGCGGCGCTTATT 60  
 QY 784 TATACATCCCAAGATCGACAGTGAAGTGGGCGGCGGCGCTTATTGTC 843  
 DB 61 TATACATCCCAAGATCGACAGTGAAGTGGGCGGCGGCGCTTATTGTC 120  
 QY 844 CGGAGTTCTGTGCGCAAGCTCACTGGGTAACAACTGCGGTGAGCAGGATAGTC 903  
 DB 121 CGGAGTTCTGTGCGCAAGCTCACTGGGTAACAACTGCGGTGAGCAGGATAGTC 180  
 QY 904 CTGCGCACTGCTCATTCATCCGATCGGTAAGATCGCTCCAAAGAGCCACTCAGTTGAGC 963  
 DB 181 CTGCGCACTGCTCATTCATCCGATCGGTAAGATCGCTCCAAAGAGCCACTCAGTTGAGC 240  
 QY 964 A-----C 965  
 DB 241 AGTAAATTCATATGCTCTATATATATGCTGATATCTTAAAGAGATGTTTACAGC 300  
 QY 966 TGAGGAATTCAGGAATTCCTTGAAGCGGTACAGCGGTGCGCTGACGAGGTGA 1025  
 DB 301 TGAGGAATTCAGGAATTCCTTGAAGCGGTACAGCGGTGCGCTGACGAGGTGA 360  
 QY 1026 GTGCGTGTCTGCTGCTGATGACACACACCTCTATCTGCGGGAACACCAAGAACAA 1085  
 DB 361 GTGCGTGTCTGCTGCTGATGACACACACCTCTATCTGCGGGAACACCAAGAACAA 420  
 QY 1086 GTGCGTGTGAGGATGACAGGCGCAGACAGTGGTCAAGCTGTAATATTCGCCGGA 1145  
 DB 421 GTGCGTGTGAGGATGACAGGCGCAGACAGTGGTCAAGCTGTAATATTCGCCGGA 480  
 QY 1146 TGTAAGCTAATTGCGTCTGCTTCAATTGACAAGTCAAGTGCCTGTGCGGAGCCAGCGA 1205

DB 481 TGTAAGCTAATTGCGTCTGCTTCAATTGACAAGTCAAGTGCCTGTGCGGAGCCAGCGA 540  
 QY 1206 TGGTCAGTACATGCCACCTTCCGGGTCATGTGCGAGGTGTTTACAGGTTGCTGCTGTC 1265  
 DB 541 TGGTCAGTACATGCCACCTTCCGGGTCATGTGCGAGGTGTTTACAGGTTGCTGCTGTC 600  
 QY 1266 GCGGACTCCGCTTGAATGTTTCCGGGAGCAAGACTCAACTCTAAA-----1313  
 DB 601 CCGGACTCCGCTTGAATGTTTCCGGGAGCAAGACTCAACTCTAAAAGGTTAGCAGCT 660  
 QY 1314 -----AGTATGGAGTGTGCAGA 1330  
 DB 661 AGTTTCAGTTTCTAGCGCTAGCTAATACATATCTCTGCTTAGTATGAGTGTGCAGA 720  
 QY 1331 CGAAGAACTGGCACAGGAGCTGCTGGACATGCGGATGAGTGTTCGAGTGGGCTGGG 1390  
 DB 721 CGAAGAACTGGCACAGGAGCTGCTGGACATGCGGATGAGTGTTCGAGTGGGCTGGG 780  
 QY 1391 GCGCGGATGCTCTAGAGTTCCTGCTGTCGTCAGGCAAGTATTAAGCT-----1442  
 DB 781 GCGCGGATGCTCTAGAGTTCCTGCTGTCGTCAGGCAAGTATTAAGCTTAAAGTTT 840  
 QY 1443 -----ATGGGCTTA 1451  
 DB 841 GAGTAACGAGGTTTCTTATACATATATACAGACTCTAATTGTATCTTTCAGATGGGCTTA 900  
 QY 1452 TTAACAATCATTAATCTGTACACGCTAAGAAATACTTAGGAATAAAGTAAACGCTCT 1511  
 DB 901 TTAACAATCATTAATCTGTACACGCTAAGAAATACTTAGGAATAAAGTAAACGCTCT 960  
 QY 1512 GAGTAAA 1518  
 DB 961 GAGTAAA 967

RESULT 6  
 AAF85075  
 ID AAF85075 standard; DNA; 1859 BP.  
 XX AAF85075;  
 AC XX  
 DT 09-JUL-2001 (first entry)  
 XX Nucleotide sequence of a human Tzap gene cDNA clone.  
 DE Tzap; T cell activation; immune response; transplant rejection;  
 KW bone marrow transplantation; rheumatoid arthritis; lupus erythematosus;  
 KW multiple sclerosis; encephalitis; vasculitis; diabetes mellitus;  
 KW pancreatitis; gastritis; thyroiditis; malignant disorder; asthma;  
 KW lepramatosi; gastritis; skin tumour; adrenal tumour; lung tumour;  
 KW wound healing; growth disorder; inflammatory disease; infectious disease;  
 ss.  
 XX Homo sapiens.  
 OS  
 PH Key Location/Qualifiers  
 FT CDS 20..1477  
 FT /\*tag= a  
 FT /product= "Tzap"  
 FT /transl\_except= "(pos: 1154..1156, aa: Xaa)"  
 FT /note= "Xaa represents an unspecified residue"  
 XX  
 PN WO200132614-A2.  
 XX 10-MAY-2001.  
 PD 30-OCT-2000; 2000WO-EP10670.  
 XX 01-NOV-1999; 99US-0162675.  
 PR 25-FEB-2000; 2000US-0185016.  
 XX (UTKU/) UTKU N.



PD 07-FEB-2001.  
XX 28-JUL-2000; 2000EP-0116126.  
XX 29-JUL-1999; 99JP-0248036.  
PR 27-AUG-1999; 99JP-0300253.  
PR 11-JAN-2000; 2000JP-0118776.  
PR 02-MAY-2000; 2000JP-0183767.  
PR 09-JUN-2000; 2000JP-0241899.  
XX (HELI-) HELIX RES INST.  
XX Ota T, Isoqai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX WPI; 2001-318749/34.  
DR  
XX Primer sets for synthesizing polynucleotides, particularly the 5602  
PT full-length cDNAs defined in the specification, and for the detection  
PT and/or diagnosis of the abnormality of the proteins encoded by the  
PT full-length cDNAs -  
XX  
PS Claim 8; SEQ ID 11399; 2537pp + CD ROM; English.  
XX The present invention describes primer sets for synthesizing 5602  
CC full-length cDNAs defined in the specification. Where a primer set  
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
CC to the complementary strand of a polynucleotide which comprises one of  
CC the 5602 nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to  
CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.  
XX  
SQ Sequence 1853 BP; 378 A; 549 C; 565 G; 361 T; 0 other;  
Query Match 30.7%; Score 477.4; DB 22; Length 1853;  
Best Local Similarity 59.8%; Pred. No. 9.2e-98;  
Matches 837; Conservative 0; Mismatches 556; Indels 6; Gaps 2;  
QY 54 GATACAGCGCGGCTCGTTTACAGCGGAGAGAGCGCGCCGCAATCGACCTGCCGCG 113  
Db 80 GCTAGTGCAGTTCAGGATGAGCGGCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 139  
QY 114 AGGAATCACTACCCAGCAATTTGGGACTGATTGCAACGCGCTGCTGAAAACGAGGAAGC 173  
Db 140 GGACATACCCCGGACAGGCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 199  
QY 174 CACTCCATATTTGTTTGTGGGCGAGGATGAGATCAAGAAGAGCGCTGAGGAGACAGTT 233  
Db 200 CTGCGCATCTGGCTTCTTGTTCACGATGCTGAGATGCTCTCTCTCTCTCTCTCTCTCTCT 259  
QY 234 GGACTTGGCGTCACTGGACACCGAAACGATGATCGATATTTGTTATCATGCCACGAGCGGT 293  
Db 260 GGAGTCCCGAGCGAGTGGACACAGAGAGGCTCTAGACATCATCTACCGCCACAGGCTAT 319  
QY 294 TTTCAAAGTCCGCCAGTGCAGAGATGCAGAGTTCATGCCGCGGACGCGCGGCTGT 353  
Db 320 CTTACAGATCCGGGCTGTGACTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 379

QY 354 GGTTCGCTGAATTTTCAGCCCGGATGGTCTCATCTCGCCAGTGGAGTGGCGACACAC 413  
Db 380 CATTTCTGTGGCTTTCAGCCCTACGGGAAAGTACCTGGCAGTGGCTCTGGAGACACAC 439  
QY 414 AGTGGATTTGGGATCTTAACACAGACACACACCTTCACTCTGCACAGTGCATAGCA 473  
Db 440 CGTGGCTTCTGGGATCTCAGCAGACAGACACACATTTTCATGCNAGGACACAGCA 499  
QY 474 GTGGGTCTGTGGGTATCTCTGGGCTCCGGATGGCAACAGGTGGCGAGCGGTGCAAGC 533  
Db 500 CTGGGCTCTTAGTATATCTCTGGTCTCCAGATGGCAAGAGCTGGGCTCAGGCTCAAGAA 559  
QY 534 GGGCTCTATATATCTCTGGGACCGGAGCGGTTCAGCAGAGGGGCGACCTTGAAGT 593  
Db 560 TGGCAGATCTCTCTGGGACCGGAGCAGCAGGGAAGCAGGTGGGCGAGACCTCGCTGG 619  
QY 594 GCACNAGAACATCACTCTCTGGGACCGGAGCTATCTATCTCGGATCGGAGTGCAG 653  
Db 620 CCACAGCAGTGTGATCAGGCTCAGCTGGGAGCCCTCCATCGACACCTCTGAGTCCG 679  
QY 654 GAACTTGTCTCCGCGAGTGGAGACGGGAGCTGCCGATTTGGAGCTTAAATTTGGGCA 713  
Db 680 CTATGTGGCAGCAGCTTCCAAAGATGGCAGTGTCTGGGATCTGGGACACAACTGCAGGCG 739  
QY 714 GTGCTTATGAACATTTGGGAGCAGACACAAATCTGTGCACAGCAGTGCAGTGGTGGAGC 773  
Db 740 CTGTGAGCGCATCTCTCCCGGACACACACAGTGGTCCCTGTCTCCGTGGGAGGGA 799  
QY 774 GGGCTTATTTATCATCTCTCCAAAGATCGCAGTGAAGATGTGGCAGCAGCTGATGG 833  
Db 800 CGGGCTCTCTACTCTGCTCTCCAGGACCGCCACCATCAAAAGTCTGGAGCTCATGACGG 859  
QY 834 AATCTTGTCCGCGACGCTTCTTGGCCAAAGCTCACTGGGTAAACAACTTGGCTGAGCAC 893  
Db 860 TGTGTGTGTGGGACTCTGCAAGGCGCAGCGCCACTGGGTGAACACCATGGCCCTCAGCAC 919  
QY 894 CGATTACGCTCTCGGACACTGCTTCATTCATTCGGTGAAGATGCTCTCAAGAGCC--A 950  
Db 920 TGACTATGCCCTGGCAGCTGGGCTCTTGAACCTGTGAGGCTCAGTTAATTCCTCAAGA 979  
QY 951 CCTCAGTTTGGACACTGAGGAATTCAGGAATCTGCTTGAAGCGCTACCGGCGGTGG 1010  
Db 980 CCTCAGGATCTTGGAGAGTGAAGGAGGAGGCTCTGAGCGGATACACCTCTGAGTGGC 1039  
QY 1011 CCTCAGGAGTGGAGTCTGCTTCTCTCTGTTCCGATGACAAACCTCTATCTGTGG-- 1068  
Db 1040 GGGCCAGGCTCAGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1099  
QY 1069 -CGGAACAACACAGAAAGTGGTGGAGCGCATCAGAGGCGCAGACAGTGGTCAACGA 1127  
Db 1100 CCACGACAGGACNAAAGCCCTCTCAGTGGATCAGAGACACACAGCTCTCATCAACCA 1159  
QY 1128 TGTGAATATTTGGCGGATGTAAGCTTAATTCGCTCTGCTTCAATTTGACAGTCAAGTGG 1187  
Db 1160 GGTGCTCTTCTCTCTGATCCCGCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1219  
QY 1188 TCTGTGGCGAGCGAGGAGTGGTCACTATGCGCCACCTTCCGGGGTTCATGTGAGGCTGT 1247  
Db 1220 GCTGTGGGATGGCAGGACGCGCAAGTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1279  
QY 1248 TTACAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1307  
Db 1280 GTACAGATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1339  
QY 1308 TCTAAAGTATGAGGATGTGCAGACGAAAGAACTGGCAGAGAGCTGCCCTGGACATGCGGA 1367  
Db 1340 ACTGAAGTGTGGGATGTGAAGGCGCCAGAAAGTGGCCATGGAGCTGCCCGGCGGCGGA 1399  
QY 1368 TCAGTGTGTGGAGTGGAGTGGGCGCGGATGCTCTGAGATGCTGCTGCTGCTGCTGCTGCTGCT 1427  
Db 1400 TGAGGTATGCTGTTGACTGGAGTCCAGATGGCCAGAGAGTGGCAGAGTGGCTGAGGAGGA 1459  
QY 1428 CAAAGTATTAAGCTTTGG 1446



Db 1329 ACTGAAGGTGGGATGTGAAGGCCAGAGACTGGCCATGGAGCTGCCGGCCACGGCGA 1388  
QY 1368 TGAGGTGTTCCGAGTGGACTGGGCGCCGAGTGGCTCTAGAGTTGCCCTGTGGCAAGGA 1427  
Db 1389 TGAGTATATCTGTTGACTGGAGTCCAGATGGCCAGAGAGTGGCAAGTGGTGGGAAGGA 1448  
QY 1428 CAAAGTTATAAGCTATGG 1446  
Db 1449 CAATGCTCCGGATGG 1467

## RESULT 9

ABN59719  
ID ABN59719 standard; cDNA; 2638 BP.

AC ABN59719;

XX

XX

DT 28-JUN-2002 (first entry)

XX

DE Novel human coding sequence SEQ ID NO: 130.

XX

KW Human; antianemic; vulnery; antiinflammatory; immunomodulator;  
KW antinfertility; cerebroprotective; cytostatic; rheumatic; gene therapy;  
KW neuroprotective; antiparkinsonian; protein therapy; EST;  
KW expressed sequence tag; gene; ss.

XX

OS Homo sapiens.

XX

XX

PN W0200222660-A2.

XX

PD 21-MAR-2002.

XX

PF 10-SEP-2001; 2001WO-US26015.

XX

PR 11-SEP-2000; 2000US-0659671.

XX

PA (HYSE-) HYSEQ INC.

XX

XX

PI Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

An isolated polynucleotide for treating diseases associated with its  
encoded polypeptide such as cancer and multiple sclerosis -

Claim 1; SEQ ID NO 130; 509pp; English.

The present invention provides the protein and coding sequences of 444  
novel human proteins. These were isolated from expressed sequences tags  
(ESTs). They can be used to stimulate cell growth, to regulate  
hematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth  
e.g. in burn treatment, to regulate the immune system e.g. to treat  
multiple sclerosis, to regulate actin or inhibit e.g. to treat  
infertility, to regulate haemostasis or thrombolysis e.g. to treat  
stroke and cancer, to screen for drugs, to treat inflammatory conditions  
e.g. rheumatoid arthritis, and to treat nervous system disorders e.g.  
Parkinson's disease. The present sequence is a coding sequence of the  
invention.

Sequence 2638 BP; 566 A; 767 C; 764 G; 541 T; 0 other;

Query Match 30.7%; Score 477.4; DB 24; Length 2638;

Best Local Similarity 59.8%; Pred. No. 1e-97;

Matches 837; Conservative 0; Mismatches 556; Indels 6; Gaps 2;

QY 54 GATACAGCGCGCTCTGTTACACGGGCGAGAGAGCGCGCCCAATCGACCTGCCGCG 113

Db 132 GCTAGTGCAGTTCAGGATGAGGCGGCGAGCTGCTGGGTTCCCGTTCCGAGTGC 191

QY 114 AGGAATCACTACCCAGCAATTTGGGACTGATTTGCAACGCGCTGCTGAAACAGGAG 173

Db 192 GGACATCACCCCGSACAGGCTCGACCTCGTGTCAACGCGCTACTTGGCCAGGAGATCC 251  
QY 174 CACTCCATATTTTGTGTTTGGGCGAGGATGAGATCAAGAGAGAGCTTGGAGGACAGCTT 233  
Db 252 COTGCGCACTGGCTTTCTTTGTTCCACGATGCTGAGATCTCTCCTCCTACTGGGGAAGCTT 311  
QY 234 GGACTTGGCTGAGTGCAGCAACGATGATGATTTGTTGATGATGATGATGATGATGATG 293  
Db 312 GGATGCCAGGAGTGGAGACAGAGAGTCTTACATCATCTACAGCCACAGGCTAT 371  
QY 294 TTTCAAGTGGCCCGAGTGCAGAGATGCAGAGTTCATGCGGGGACACGCGGAGGCTGT 353  
Db 372 CTTTCAAGTCCGGGCTGTGACTCGCTGCACGAGCTCTTGGAGGCTTCAAGTGCAGGCT 431  
QY 354 GGTTTGGCTGAATTTTTCAGGCGGATGCTCTCTCTCCAGTGGAGTGGCGACACAC 413  
Db 432 CATTTCTGTGCTTTCAGCCCTACGCGAAAGTACCTTGGCAGTGGCTCTGGAGACAC 491  
QY 414 AGTGGATTTGGGATCTTAAACACAGACACACGAGTTCACCTGCACAGTCTAAGCA 473  
Db 492 CGTGGCTTCTGGGATCTCAGACAGACACACACATTTTACATGCAAGGACACAGACA 551  
QY 474 GTGGTCTTGTGATCTCTCTGGGCTCCGAGTGGCAACGCTTGGCGAGGCTTGCAGG 533  
Db 552 CTGGTCTCTAGTATATCTCTCTCCAGATGGCAAGAGTGGCTTGGCTTCAAGAA 611  
QY 534 GGGCTCTATATCATCTGGGACCGGAGCGGCTCAGCAGAGGGGCGACCTTGGTGG 593  
Db 612 TGGCAGATTTCTCTTGGGACCAAGCAGAGGAGAGTGGGCGAGGACCTCGCTGG 671  
QY 594 GCACAGAAACATCACTGCTCGCTGGGAACTGATATCATCTGGGAGTGGGAGTGGAG 653  
Db 672 CCACAGCAAGTGTGATCAGAGGCTGAGTGGGAGCCCTCCATGCGAAGCTTGAAGTGC 731  
QY 654 GAACTTGTCTTCCGCGAGTGGAGCGGAGTGGGAGTGGGAGTGGGAGTGGGAGTGG 713  
Db 732 CTATGTGCGCAGCTTCCAGGATGGCAGTGTGCGGATCTGGGACACAACTGCGAGCG 791  
QY 714 GTGCTTATGACATTTGCCGAGACACAAATGCTGTGACAGAGTGGAGTGGGAGTGG 773  
Db 792 CTGTGAGGAGTCTCTCAGCGGACACCCAGTGGTCTCTCTCTCTCTCTCTCTCTCT 851  
QY 774 GGGCTTATTTATACATCTCTCAAGATGCGACAGTGAAGTGGCGAGGAGTGGAGTGG 833  
Db 852 CGGCTTCT 911  
QY 834 AATCTTGTGCGGAGCTTCTCTGCGCAAGCTCTCTGCGCAAGCTCTCTGCGCTGAG 893  
Db 912 TGTGCTGTGCGGAGTCTCTCAAGGCGCGGAGTGGGAGTGGGAGTGGGAGTGGGAG 971  
QY 894 CGATTAGCTCTCTGCGCAGTGGTCCATTCATCGGTTGAAGGATCGCTCCAGAGCC--A 950  
Db 972 TGACTATGCTCTGCGCAGTGGGCGCTTTGAACCTGTGAGGCTCAGTTATTCCTCAAGA 1031  
QY 951 COTCAGTTGAGCAGTGGAGAAATTCGAGAAATTCGCTTGAAGCGTACAGCGCGTGTG 1010  
Db 1032 COTCCAGAGTCTCTTCAAGAGTGTGAAGGAGAGGCTCTGAGCGGATACACCTCTG 1091  
QY 1011 CCCTGAGCAGTGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1068  
Db 1092 GGGCCAGGCTCCAGAGAGGCTGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1151  
QY 1069 -CGGAAACACCAACAAAGTGGTGGAGCGTGGAGCGGACAGGCGGAGGCTGCAAGCA 1127  
Db 1152 CCCAGAGAGGACAAAGAGCTCTCCTCTGAGTACAGAGACACAGAGCTCTCATCACCA 1211  
QY 1128 TGTGAATATTCGCCGAGTGAAGCTTAATTTGCGTCTGCTTCTATTTGAAAGTCA 1187  
Db 1212 GTGTGCTTCT 1271  
QY 1188 TCTGTGCGAGGAGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1247  
Db 1272 GCTGTGGATGGCAGAGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1331







PR	21-JUL-1999;	99US-0144814.	PR	18-OCT-1999;	99US-0159584.
PR	21-JUL-1999;	99US-0145086.	PR	21-OCT-1999;	99US-0160741.
PR	21-JUL-1999;	99US-0145086.	PR	21-OCT-1999;	99US-0160767.
PR	22-JUL-1999;	99US-0145085.	PR	21-OCT-1999;	99US-0160768.
PR	22-JUL-1999;	99US-0145087.	PR	21-OCT-1999;	99US-0160770.
PR	22-JUL-1999;	99US-0145089.	PR	21-OCT-1999;	99US-0160814.
PR	22-JUL-1999;	99US-0145192.	PR	21-OCT-1999;	99US-0160815.
PR	23-JUL-1999;	99US-0145145.	PR	21-OCT-1999;	99US-0160815.
PR	23-JUL-1999;	99US-0145218.	PR	22-OCT-1999;	99US-0160980.
PR	23-JUL-1999;	99US-0145224.	PR	22-OCT-1999;	99US-0160981.
PR	26-JUL-1999;	99US-0145276.	PR	22-OCT-1999;	99US-0160989.
PR	27-JUL-1999;	99US-0145913.	PR	25-OCT-1999;	99US-0161404.
PR	27-JUL-1999;	99US-0145918.	PR	25-OCT-1999;	99US-0161405.
PR	27-JUL-1999;	99US-0145919.	PR	25-OCT-1999;	99US-0161406.
PR	28-JUL-1999;	99US-0145951.	PR	26-OCT-1999;	99US-0161359.
PR	02-AUG-1999;	99US-0146386.	PR	26-OCT-1999;	99US-0161360.
PR	02-AUG-1999;	99US-0146388.	PR	26-OCT-1999;	99US-0161361.
PR	02-AUG-1999;	99US-0146389.	PR	28-OCT-1999;	99US-0161920.
PR	03-AUG-1999;	99US-0146389.	PR	28-OCT-1999;	99US-0161920.
PR	04-AUG-1999;	99US-0147038.	PR	28-OCT-1999;	99US-0161992.
PR	04-AUG-1999;	99US-0147204.	PR	28-OCT-1999;	99US-0161993.
PR	04-AUG-1999;	99US-0147302.	PR	29-OCT-1999;	99US-0162142.
PR	05-AUG-1999;	99US-0147302.	Query Match 21.4%; Score 333.4; DB 21; Length 1726;		
PR	05-AUG-1999;	99US-0147192.	Best Local Similarity 54.4%; Pred. No. 2.2e-65;		
PR	05-AUG-1999;	99US-0147260.	Matches 723; Conservative		
PR	06-AUG-1999;	99US-0147303.	Mismatches 591; Indels 15; Gaps 2;		
PR	06-AUG-1999;	99US-0147416.			
PR	09-AUG-1999;	99US-0147493.			
PR	09-AUG-1999;	99US-0147935.			
PR	10-AUG-1999;	99US-0148171.			
PR	11-AUG-1999;	99US-0148319.			
PR	12-AUG-1999;	99US-0148341.			
PR	13-AUG-1999;	99US-0148565.			
PR	13-AUG-1999;	99US-0148684.			
PR	16-AUG-1999;	99US-0149368.			
PR	17-AUG-1999;	99US-0149317.			
PR	18-AUG-1999;	99US-0149426.			
PR	20-AUG-1999;	99US-0149722.			
PR	20-AUG-1999;	99US-0149723.			
PR	20-AUG-1999;	99US-0149929.			
PR	23-AUG-1999;	99US-0149902.			
PR	23-AUG-1999;	99US-0149930.			
PR	25-AUG-1999;	99US-0150566.			
PR	26-AUG-1999;	99US-0150884.			
PR	27-AUG-1999;	99US-0151065.			
PR	27-AUG-1999;	99US-0151066.			
PR	27-AUG-1999;	99US-0151080.			
PR	30-AUG-1999;	99US-0151303.			
PR	01-SEP-1999;	99US-0151438.			
PR	01-SEP-1999;	99US-0151930.			
PR	07-SEP-1999;	99US-0152363.			
PR	10-SEP-1999;	99US-0153070.			
PR	13-SEP-1999;	99US-0153758.			
PR	15-SEP-1999;	99US-0154018.			
PR	16-SEP-1999;	99US-0154039.			
PR	20-SEP-1999;	99US-0154779.			
PR	22-SEP-1999;	99US-0155139.			
PR	23-SEP-1999;	99US-0155486.			
PR	24-SEP-1999;	99US-0155659.			
PR	28-SEP-1999;	99US-0156458.			
PR	29-SEP-1999;	99US-0156596.			
PR	04-OCT-1999;	99US-0157117.			
PR	05-OCT-1999;	99US-0157753.			
PR	06-OCT-1999;	99US-0157865.			
PR	07-OCT-1999;	99US-0158029.			
PR	08-OCT-1999;	99US-0158232.			
PR	12-OCT-1999;	99US-0158369.			
PR	13-OCT-1999;	99US-0159293.			
PR	13-OCT-1999;	99US-0159294.			
PR	13-OCT-1999;	99US-0159295.			
PR	13-OCT-1999;	99US-0159329.			
PR	14-OCT-1999;	99US-0159330.			
PR	14-OCT-1999;	99US-0159331.			
PR	14-OCT-1999;	99US-0159637.			
PR	14-OCT-1999;	99US-0159638.			
PR	21-JUL-1999;	99US-0144814.	PR	18-OCT-1999;	99US-0159584.
PR	21-JUL-1999;	99US-0145086.	PR	21-OCT-1999;	99US-0160741.
PR	21-JUL-1999;	99US-0145086.	PR	21-OCT-1999;	99US-0160767.
PR	22-JUL-1999;	99US-0145085.	PR	21-OCT-1999;	99US-0160768.
PR	22-JUL-1999;	99US-0145087.	PR	21-OCT-1999;	99US-0160770.
PR	22-JUL-1999;	99US-0145089.	PR	21-OCT-1999;	99US-0160814.
PR	22-JUL-1999;	99US-0145192.	PR	21-OCT-1999;	99US-0160815.
PR	23-JUL-1999;	99US-0145145.	PR	21-OCT-1999;	99US-0160815.
PR	23-JUL-1999;	99US-0145218.	PR	22-OCT-1999;	99US-0160980.
PR	23-JUL-1999;	99US-0145224.	PR	22-OCT-1999;	99US-0160981.
PR	26-JUL-1999;	99US-0145276.	PR	22-OCT-1999;	99US-0160989.
PR	27-JUL-1999;	99US-0145913.	PR	25-OCT-1999;	99US-0161404.
PR	27-JUL-1999;	99US-0145918.	PR	25-OCT-1999;	99US-0161405.
PR	27-JUL-1999;	99US-0145919.	PR	25-OCT-1999;	99US-0161406.
PR	28-JUL-1999;	99US-0145951.	PR	26-OCT-1999;	99US-0161359.
PR	02-AUG-1999;	99US-0146386.	PR	26-OCT-1999;	99US-0161360.
PR	02-AUG-1999;	99US-0146388.	PR	26-OCT-1999;	99US-0161361.
PR	02-AUG-1999;	99US-0146389.	PR	28-OCT-1999;	99US-0161920.
PR	03-AUG-1999;	99US-0146389.	PR	28-OCT-1999;	99US-0161920.
PR	04-AUG-1999;	99US-0147038.	PR	28-OCT-1999;	99US-0161992.
PR	04-AUG-1999;	99US-0147204.	PR	28-OCT-1999;	99US-0161993.
PR	04-AUG-1999;	99US-0147302.	PR	29-OCT-1999;	99US-0162142.
PR	05-AUG-1999;	99US-0147302.	Query Match 21.4%; Score 333.4; DB 21; Length 1726;		
PR	05-AUG-1999;	99US-0147192.	Best Local Similarity 54.4%; Pred. No. 2.2e-65;		
PR	05-AUG-1999;	99US-0147260.	Matches 723; Conservative		
PR	06-AUG-1999;	99US-0147303.	Mismatches 591; Indels 15; Gaps 2;		
PR	06-AUG-1999;	99US-0147416.			
PR	09-AUG-1999;	99US-0147493.			
PR	09-AUG-1999;	99US-0147935.			
PR	10-AUG-1999;	99US-0148171.			
PR	11-AUG-1999;	99US-0148319.			
PR	12-AUG-1999;	99US-0148341.			
PR	13-AUG-1999;	99US-0148565.			
PR	13-AUG-1999;	99US-0148684.			
PR	16-AUG-1999;	99US-0149368.			
PR	17-AUG-1999;	99US-0149317.			
PR	18-AUG-1999;	99US-0149426.			
PR	20-AUG-1999;	99US-0149722.			
PR	20-AUG-1999;	99US-0149723.			
PR	20-AUG-1999;	99US-0149929.			
PR	23-AUG-1999;	99US-0149902.			
PR	23-AUG-1999;	99US-0149930.			
PR	25-AUG-1999;	99US-0150566.			
PR	26-AUG-1999;	99US-0150884.			
PR	27-AUG-1999;	99US-0151065.			
PR	27-AUG-1999;	99US-0151066.			
PR	27-AUG-1999;	99US-0151080.			
PR	30-AUG-1999;	99US-0151303.			
PR	01-SEP-1999;	99US-0151438.			
PR	01-SEP-1999;	99US-0151930.			
PR	07-SEP-1999;	99US-0152363.			
PR	10-SEP-1999;	99US-0153070.			
PR	13-SEP-1999;	99US-0153758.			
PR	15-SEP-1999;	99US-0154018.			
PR	16-SEP-1999;	99US-0154039.			
PR	20-SEP-1999;	99US-0154779.			
PR	22-SEP-1999;	99US-0155139.			
PR	23-SEP-1999;	99US-0155486.			
PR	24-SEP-1999;	99US-0155659.			
PR	28-SEP-1999;	99US-0156458.			
PR	29-SEP-1999;	99US-0156596.			
PR	04-OCT-1999;	99US-0157117.			
PR	05-OCT-1999;	99US-0157753.			
PR	06-OCT-1999;	99US-0157865.			
PR	07-OCT-1999;	99US-0158029.			
PR	08-OCT-1999;	99US-0158232.			
PR	12-OCT-1999;	99US-0158369.			
PR	13-OCT-1999;	99US-0159293.			
PR	13-OCT-1999;	99US-0159294.			
PR	13-OCT-1999;	99US-0159295.			
PR	13-OCT-1999;	99US-0159329.			
PR	14-OCT-1999;	99US-0159330.			
PR	14-OCT-1999;	99US-0159331.			
PR	14-OCT-1999;	99US-0159637.			
PR	14-OCT-1999;	99US-0159638.			







||||| 361 TGGTCTCGGTCGAGCACTTACCTTATCTGTCGCCAGCAGAGCAAAAGC 420  
1088 GGGTTGAGCGATGACAGAGGACAGAGAGTGGTCAACAGATGTAATAATATGCCGAGTG 1147  
421 CTCTCACTCGGATGACAGGACACCAAGCTCTCATCAACAGTGCTCTTCTCTCTGACT 480  
1148 TAAAGTAAATGCTGCTGCTTATTTGACAAGTCAAGTCTGCTGCGGAGCAGGATG 1207  
481 CCGCATGCTGGCTAGTGCCTCTTTGACAAGTCAAGCTGTGGATGGCAGAGCG 540  
1208 GTCACTACATGCGCACCTTCCGGGTCATGTCAGAGCTGTTTACACGGTGTGCTGTCGG 1267  
541 GCAAGTACCTGGCTTCCCTACGCGGCCAGCTGGCTGTACCAAGTGGCTGTGCTCAG 600  
1268 CGGACTCCGCTTGNATTTTCCGCGCAGCAAGACTCAACTTAAAGTATGAGTGC 1327  
601 CTGACAGTGGCTCTGCTGAGCGGCGAGCAGTGACAGCACACTGAAGTGTGGATGTA 660  
1328 AGACGAAGAACTGGCAGAGAGTCTGCTGGACATGCGGATGAGTGTTCGGAGTGGACT 1387  
661 AGCCCCAGAAGCTGGCCATGAGCTGCGCGGCCAGCGGATGAGGTATATGCTGTGACT 720  
1388 GGGCGCCGAGTGGCTAGAGTTGCTCTGTTGGCAAGCAAAAGTTATAAGCTATGG 1446  
721 GGAGTCCAGATGGCCAGAGAGTGGCAAGTGGTGGGAGGACAAATGCTCCGATATGG 779

RESULT 14  
AAH30571  
ID AAH30571 standard; cdna; 399 BP.  
XX  
AC AAH30571;  
XX  
DT 27-JUL-2001 (first entry)  
XX  
DE Human colon cancer cell line Km12L4-A cdna library derived sequence #505.  
XX  
KW Human; diagnosis; colon cancer; cancer; malignant; chromosome mapping;  
KW detection; colon cancer cell line Km12L4-A; ss.  
XX  
OS Homo sapiens.  
XX  
PN W0200018916-A2.  
XX  
PD 06-APR-2000.  
XX  
PF 23-SEP-1999; 99WO-US22226.  
XX  
PR 28-SEP-1998; 98US-0102161.  
PR 28-SEP-1998; 98US-0102180.  
PR 29-SEP-1998; 98US-0102380.  
PR 08-OCT-1998; 98US-0103815.  
PR 27-OCT-1998; 98US-0105877.  
XX  
PA (CHIR ) CHIRON CORP.  
PA (HYSE-) HYSEQ INC.  
XX  
PI Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;  
PI Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;  
PI Lanson G, Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;  
PI Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;  
XX  
DR WPI; 2000-293155/25.  
XX  
PT Polynucleotide library comprising 1079 defined sequences, useful in  
PT the form of an array to detect cancer or susceptibility to cancer -  
XX  
PS Claim 1; Page 328-329; 502pp; English.  
XX  
CC The present invention describes a library of polynucleotides comprising  
CC 1079 nucleotide sequences (given in AAH30067 to AAH31145). Also described  
CC are: (1) an isolated polynucleotide (1) having at least 90% identity to

one of the 1079 sequences; (2) a recombinant host cell containing (1);  
(3) an isolated polypeptide (11) encoded by (1); (4) an antibody that  
specifically binds to (11); (5) a vector comprising (1); and (6) a method  
of detecting differentially expressed genes correlated with a cancerous  
state of a mammalian cell comprising detecting a gene product encoded by  
one of the 1079 sequences given in the specification. The polynucleotides  
are used to monitor patients having (or susceptible) to cancer to detect  
potentially malignant events at a molecular level before they are  
detectable at a gross morphological level. The polynucleotides are also  
useful for monitoring the efficacy of various therapies and preventive  
interventions. Polynucleotide probes based on the disclosed sequences  
are useful for chromosome mapping and detection of transcription levels.  
The 1079 polynucleotide sequences were derived from a human colon cancer  
cell line Km12L4-A cdna library.  
XX  
SQ Sequence 399 BP; 92 A; 120 C; 112 G; 75 T; 0 other;  
Query Match 11.3%; Score 176; DB 21; Length 399;  
Best Local Similarity 66.1%; Pred. No. 3.9e-30;  
Matches 254; Conservative 0; Mismatches 130; Indels 0; Gaps 0;  
QY 266 TCGATATTGTATACAGCCACAGCGGCTTTTCAAGTGCAGTGCACAGATGCACGA 325  
DB 16 TAGACATCATCTACAGCCACAGGCTATCTTCAGAGTCCGGCTGTGACTCGCTGCACCA 75  
QY 326 GTTCCATGCGCGCACAGCGGAGCGCTGTTCGCTGAATTTTCAGCCGGATGTGCTC 385  
DB 76 GCTCTTGGAGGTACAGTGAAGCAGTATTTCTGTGGCTTCAGCCCTACGGAAAGT 135  
QY 386 ATCTCGCAGTGAAGTGGCAGACACACAGTGCAGTGTGGATCTTAACACAGACAC 445  
DB 136 ACCTGGCGATGCTCTGGAGACACACACCGCTCTCTGGGATCTCAGCAGACAGAC 195  
QY 446 CGCACTTCACCTGCACAGGTCAATAGCAGTGGTTCGCTGATCTCGGCTCCGGATG 505  
DB 196 CACATTTTCATGCAAGGACACACAGTGGTCTTATATATCTGTCTCCAGATG 255  
QY 506 GCAACCGTGTGGCCAGCGGTGCAAGCGGCTCTATATATCTCTGGACCCGAGACGG 565  
DB 256 GCAAGAAGCTGGCTCAGGCTGCAAGAATGGCCAGATTCCTCTGGGACCAACGACAG 315  
QY 566 GTCACAGAAAGGCGACCGCTTGAAGTGGGACAAAGAACACATCAATCGCTCGCTGGG 625  
DB 316 GGAAGCAGGTGGCAGGACCGCTCGCTGGCCACAGCAAGTGGATCAGAGCCTGAGCTGG 375  
QY 626 AACGATATCATCCGATCCGGAGT 649  
DB 376 AGCCCTCCATGCAAGCCTCTGAGT 399  
RESULT 15  
AAH05727  
ID AAH05727 standard; cdna; 547 BP.  
XX  
AC AAH05727;  
XX  
DT 26-JUN-2001 (first entry)  
XX  
DE Human cdna clone (5'-primer) SEQ ID NO:2562.  
XX  
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
XX  
OS Homo sapiens.  
XX  
PN EP1074617-A2.  
XX  
PD 07-FEB-2001.  
XX  
PF 28-JUL-2000; 2000EP-0116126.  
XX  
PR 29-JUL-1999; 99JP-0248036.  
PR 27-AUG-1999; 99JP-0300253.  
PR 11-JAN-2000; 2000JP-0118776.

PR

02-MAY-2000; 2000JP-0183767.

PR

09-JUN-2000; 2000JP-0241899.

XX

(HELI-) HELIX RES INST.

XX

PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI

Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX

WPI; 2001-318749/34.

XX

Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs -

XX

Claim 1; SEQ ID 2562; 2537pp + CD ROM; English.

XX

The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.

XX

SQ Sequence 547 BP; 108 A; 159 C; 168 G; 109 T; 3 other;

## Query Match

10.5%; Score 164; DB 22; Length 547;

Best Local Similarity 62.0%; Pred. No. 2.1e-27;

Matches 274; Conservative 0; Mismatches 167; Indels 1; Gaps 1;

QY

54 GATACAGCGCGCTCGTTTACACGGCGAGGAGCGCGCCGCAATCGACCTGCGGC 113

DB

80 GCTAGTCAGTCCAGATGAGGGCGGCACCTGCTGGTTCCTCCGTCGACGTCCCGT 139

QY

114 AGGAATCACTACCCAGCAATTGGGACTGATTGTCAGCGCTGCTGAAAACGAGGAGC 173

DB

140 GGACATCACCCCGGACACAGGCTGCTGTGTCAGCGCTACTGGCCCGAGGAGATCC 199

QY

174 CACTCCCATATTTGTTTGTGGGAGGATGATCAAGAGCGCTGGAGACAGCTT 233

DB

200 CTGCGACTGGCTTTCTTTGTCCAGATGCTGAGATGCTCTCTACTGGGAGAGAGCTT 259

QY

234 GGACTTGGCTCAGTGGACACCGAAACGTGATCGATATTGTATFACGCCACAGGGCGT 293

DB

260 GGAGTCCCGGCGAGTGGAGACAGAGAGGTCCTAGACATCATCTACCCAGCCAGGCTAT 319

QY

284 TTTCAAGTGGCGCCAGTGACAAAGATGACGAGTTCCTCCGCGGACACGCGGAGGCTGT 353

DB

320 CTTTCCAGTCCGGCTGTGACTCGCTGACCCAGCTCTTTGGAGGGTTCACGTGAGGCAGT 379

QY

354 GTTTCCTGAATTTACGCCGGATGGTGTCTATCTCGCCAGTGGAGTGGGACACAC 413

DB

380 CATTTCGTGGCTTCAGCCCTACGGGAAAGTACCTGGCCAGTGGCTCTGGAGACAC 439

QY

414 AGTCGATTTGGGATCTTAACACAGACACACCGACTTCACCTGCACAGGTTCATAAGCA 473

DB

440 CGTGGCTTCTGGGATCTCAGCACAGACAGACACACATTTTCACATGCA-AGGACACNGACA 498

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 2, 2003, 03:06:49 ; Search time 101 Seconds  
(without alignments)  
4721.608 Million cell updates/sec

Title: US-09-830-980-2

Perfect score: 1555

Sequence: 1 aattcccaaaatgcagga.....aaaaaaaaaaaaaaaaaaaa 1555

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_NA:\*

1: /cgn2.6/ptodata/1/ina/5A\_COMB.seq:\*

2: /cgn2.6/ptodata/1/ina/5B\_COMB.seq:\*

3: /cgn2.6/ptodata/1/ina/6A\_COMB.seq:\*

4: /cgn2.6/ptodata/1/ina/6B\_COMB.seq:\*

5: /cgn2.6/ptodata/1/ina/PCTUS\_COMB.seq:\*

6: /cgn2.6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result		Query		SUMMARIES		Description	
No.	Score	Match	Length	ID			
1	60.2	3.9	2359	1	US-08-188-582-4	Sequence 4, Appl	
2	60.2	3.9	2359	1	US-08-646-715-4	Sequence 4, Appl	
3	59	3.8	2674	4	US-09-817-180-1	Sequence 1, Appl	
4	57.6	3.7	2250	6	5258287-23	Patent No. 5258287	
5	57.4	3.7	1221	3	US-08-965-600-2	Sequence 2, Appl	
6	57.4	3.7	1221	4	US-09-489-506-2	Sequence 2, Appl	
7	56.2	3.6	10136	1	US-08-353-700-2	Sequence 2, Appl	
8	56.2	3.6	10136	5	PCT-US95-16216-2	Sequence 2, Appl	
9	56	3.6	1872	4	US-09-801-052-1	Sequence 1, Appl	
10	55.8	3.6	2085	2	US-08-283-917-8	Sequence 8, Appl	
11	55.8	3.6	2085	2	US-08-961-716-8	Sequence 8, Appl	
12	55.4	3.6	1736	3	US-09-182-816-22	Sequence 22, Appl	
13	55.4	3.6	1736	3	US-09-182-816-24	Sequence 24, Appl	
14	55.4	3.6	1736	3	US-09-471-528-22	Sequence 22, Appl	
15	55.4	3.6	1736	3	US-09-471-528-24	Sequence 24, Appl	
16	55.4	3.6	1736	4	US-09-634-530-22	Sequence 22, Appl	
17	55.4	3.6	1736	4	US-09-634-530-24	Sequence 24, Appl	
18	55.4	3.6	7218	1	US-08-232-463-14	Sequence 14, Appl	
19	55.2	3.5	812	4	US-09-091-097-7	Sequence 7, Appl	
20	54.8	3.5	1212	4	US-09-182-145-34	Sequence 34, Appl	
21	54.8	3.5	1212	4	US-09-182-145-35	Sequence 35, Appl	
22	54.6	3.5	2323	4	US-09-149-476-24	Sequence 24, Appl	
23	54.4	3.5	1098	3	US-09-248-335-35	Sequence 35, Appl	
24	54.2	3.5	578	4	US-09-602-877A-95	Sequence 95, Appl	
25	54.2	3.5	1066	1	US-08-157-101A-4	Sequence 4, Appl	
26	53.8	3.5	1117	4	US-09-247-373B-33	Sequence 33, Appl	
27	53.6	3.4	6671	1	US-08-280-443-1	Sequence 1, Appl	

Sequence 1, Appl  
Sequence 1, Appl  
Sequence 1, Appl  
Sequence 1, Appl  
Sequence 1, Appl  
Sequence 1, Appl  
Sequence 14, Appl  
Sequence 5, Appl  
Sequence 5, Appl  
Sequence 10, Appl  
Sequence 12, Appl  
Sequence 151, Appl  
Sequence 26, Appl  
Sequence 208, Appl  
Sequence 1, Appl  
Sequence 64, Appl  
Sequence 39, Appl  
Sequence 59, Appl  
Sequence 6, Appl

28 53.6 3.4 6671 1 US-08-457-459-1  
29 53.6 3.4 6671 1 US-08-555-678-1  
30 53.6 3.4 6671 5 PCT-US95-02275-1  
31 53.4 3.4 2184 4 US-08-955-918C-1  
32 53.4 3.4 2184 4 US-08-697-766A-1  
33 53.4 3.4 2447 2 US-09-014-969-14  
34 53.2 3.4 1411 4 US-08-964-127-5  
35 53.2 3.4 1411 4 US-09-496-692-5  
36 53.2 3.4 1582 3 US-08-545-136B-10  
37 53.2 3.4 1582 3 US-09-370-838-12  
38 53.2 3.4 3275 4 US-08-702-344-26  
39 53 3.4 144 1 US-08-702-344-26  
40 53 3.4 1378 4 US-09-149-476-208  
41 52.8 3.4 3124 4 US-09-734-030-1  
42 52.6 3.4 1474 4 US-08-821-994-64  
43 52.6 3.4 1878 4 US-09-465-558-39  
44 52.6 3.4 2202 4 US-09-465-558-59  
45 52.2 3.4 240 1 US-08-628-417-6

# ALIGNMENTS

## RESULT 1

US-08-188-582-4  
; Sequence 4, Application US/08188582  
; Patent No. 5534410  
; GENERAL INFORMATION:  
; APPLICANT: Tjian, Robert  
; APPLICANT: Comai, Lucio  
; APPLICANT: Dynlacht, Brian D.  
; APPLICANT: Hoey, Timothy  
; APPLICANT: Ruppert, Siegfried  
; APPLICANT: Tanese, Naoko  
; APPLICANT: Wang, Edit  
; APPLICANT: Weinzierl, Robert O.J.  
; TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,  
; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAPS AND METHODS OF USE  
; NUMBER OF SEQUENCES: 36  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT  
; STREET: 4 Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/188,582  
; FILING DATE: 28-JAN-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Osman, Richard A  
; REGISTRATION NUMBER: 36,627  
; REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 781-1989  
; TELEFAX: (415) 398-3249  
; TELEX: 910 277299  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2359 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 49..2160





; LENGTH: 2674  
; TYPE: DNA  
; ORGANISM: Human  
US-09-817-180-1

Query Match  
Best Local Similarity 3.8%; Score 59; DB 4; Length 2674;  
Matches 116; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

QY 1345 CAGGAGCTGCTGGACATGGGATGAGGTTCGGAGTGGAGTGGCGCCGCGATGGCTCT 1404  
DB 2439 CAGCATCCACTGCCGCGAGGATGAGCGCGGTCTCTGTGCTGCTGCTGCTGCC 2498  
QY 1405 AGAGTTGCTCTGGTGGCAAGCAAGATTATAAGCTATGGGCTATTAAACAAATCAT 1464  
DB 2499 AGGCTTCTCTTCGGGCGCAACATATAAACCACTTGTGCCCACTGAAAAA 2558  
QY 1465 AACTTGTACACGTAAGAAATACCTAGGAATAAAGTAAACGCTCCTGAGTAAAAA 1524  
DB 2559 AAAAAA 2618  
QY 1525 AAAAAA 1555  
DB 2619 AAAAAA 2649

RESULT 4  
5258287-23  
; Patent No. 5258287  
; APPLICANT: BAXTER, ROBERT C. WOOD, WILLIAM I.  
; TITLE OF INVENTION: DNA ENCODING AND METHODS OF PRODUCTION  
; OF INSULIN-LIKE GROWTH FACTOR BINDING PROTEIN BP53  
; NUMBER OF SEQUENCES: 58  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/171,623  
; FILING DATE: 22-MAR-1988  
; SEQ ID NO: 23:  
; LENGTH: 2550  
5258287-23

Query Match  
Best Local Similarity 3.7%; Score 57.6; DB 6; Length 2550;  
Matches 84; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 1426 GACAAAGTTATAAGCTATGGCTTTATAACAAATCAATTAAGTACACGTAAGAAA 1485  
DB 2423 GAGAAGTTGCTTGCATGTATTATATAATAGTAAATTAAGTTTACCATTA 2482  
QY 1486 TACTTAGGAATAAGTAAACGCTCCTGAGTAAAAA 1545  
DB 2483 AAAAAA 2542  
QY 1546 AAAAAA 1553  
DB 2543 AAAAAA 2550

RESULT 5  
US-08-965-600-2  
; Sequence 2, Application US/08965600  
; Patent No. 6077688  
; GENERAL INFORMATION:  
; APPLICANT: Bandman, Olga  
; APPLICANT: Lal, Preeti  
; APPLICANT: Corley, Neil C.  
; APPLICANT: Shah, Purvi  
; TITLE OF INVENTION: NEW TRANSDUCIN BETA-1 SUBUNIT  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA

; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/965,600  
; FILING DATE: Herewith  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0416 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-855-0555  
; TELEFAX: 650-845-4166  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1221 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: KIDN0T02  
; CLONE: 194046  
; US-08-965-600-2

Query Match  
Best Local Similarity 3.7%; Score 57.4; DB 3; Length 1221;  
Matches 157; Conservative 0; Mismatches 166; Indels 0; Gaps 0;

QY 1130 TGAATATTCCGCGGATGTAAGCTAATTTGGCTGCTGCTTCAATTTGACAGTCAAGTGC 1189  
DB 666 TGACCTTTTCCCGGACTCCAGCTCTTGTCACTGCTTCAGATGATGGCTACATCAAGA 725  
QY 1190 TGTGGCGAGCAGCGATGGTGCAGTACATGGCCACCTTCCGGGGTGCATGTGAGGCTGTT 1249  
DB 726 TCTATGATGATACAACTGCAATTTGGCTGGCAGCTGAGCGCCATGCTCTGGGTGC 785  
QY 1250 ACAGGTTGCTGGTCCGCGGACTCCCGCTTGATTTTCCGCGCAGCAAGACTCAACTC 1309  
DB 786 TGAAGTTGCAATTTCTGCTGATGACACTCACTTTGTTCCAGTTGCTGTGACAAAATG 845  
QY 1310 TAAAGTATGGAGTGTGCAGACGAAAGAACTGGCAGGAGCTGCTGGACATGCGGATG 1369  
DB 846 TAAAGTTTGGGATGTTGGAACGAGGACTTGTGTTACACCTTCTTTGATCACCAGGATC 905  
QY 1370 AGGTGTCGAGTGGAGTGGCGGCGGATGGCTAGAGTTGCTCTAGAGTTGCTCTGTGCAAGGACA 1429  
DB 906 AGGTGTCGAGTGGAGTGGCGGCGGATGGCTAGAGTTGCTCTAGAGTTGCTCTGTGCAAGGACA 1429  
QY 1430 AAGTTATAAGCTATGGGCTTAT 1452  
DB 966 AGGAATTCACATCTATGATTCT 988

RESULT 6  
US-09-489-506-2  
; Sequence 2, Application US/09489506  
; Patent No. 6465619  
; GENERAL INFORMATION:  
; APPLICANT: Bandman, Olga  
; APPLICANT: Lal, Preeti  
; APPLICANT: Corley, Neil C.  
; APPLICANT: Shah, Purvi  
; TITLE OF INVENTION: NEW TRANSDUCIN BETA-1 SUBUNIT  
; NUMBER OF SEQUENCES: 3

; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/489,506  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/965,600  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0416 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-855-0555  
; TELEFAX: 650-845-4166  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1221 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: KIDNOT02  
; CLONE: 194046  
; US-09-489-506-2

Query Match 3.7%; Score 57.4; DB 4; Length 1221;  
Best Local Similarity 48.6%; Pred. No. 2.6e-06;  
Matches 157; Conservative 0; Mismatches 166; Indels 0; Gaps 0;  
Qy 1130 TGAATATTCCCGGATGTAAGCTAATTGGCTGCTGCTTCATTGGACAAAGTCAGTGCCTC 1189  
Db 666 TGACCTTTTCCCGGACTCCAGCTCTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 725  
Qy 1190 TGTGGCGAGCCAGCGATGTCAGTACATGCGCCACCTTCGCGGGTCAATGTCAGGCTGTTT 1249  
Db 726 TCTATGATGTACACATGCCAATTTGGCTGGCAGCGCTGAGCGGCCATGCTCCTGGGTGC 785  
Qy 1250 ACACGGTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1309  
Db 786 TGAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 845  
Qy 1310 TAAAGTATGAGTGTGCAGACGAGAACTGGCACAGGAGCTGCCCTGGACATGGGATG 1369  
Db 846 TAAAGTGTGAGTGTGGAGCAGGAGCTGTGTTCACACCTTCTTTGATCACCAGGATC 905  
Qy 1370 AGGTGTTCCGAGTGGAGTGGCGGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1429  
Db 906 AGGTGTTGGAGTAAATAACAATGGAATGTTCAAAAATGTTGCTGCTGTTGGAGATGACC 965  
Qy 1430 RAGTATTAAGCTATGGCTTAT 1452  
Db 966 AGGAATTCACATCATGATGTTG 988

RESULT 7  
US-08-353-700-2  
; Sequence 2, Application US/08353700  
; Patent No. 5599919  
; GENERAL INFORMATION:  
; APPLICANT: YEN, TIMOTHY J.

; APPLICANT: RATTNER, JEROME B.  
; TITLE OF INVENTION: NUCLEIC ACID ENCODING A  
; TITLE OF INVENTION: TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN,  
; TITLE OF INVENTION: AND METHODS OF USE  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: DANN, DORFMAN, HERRELL AND SKILLMAN  
; STREET: 1601 MARKET STREET, SUITE 720  
; CITY: PHILADELPHIA  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19103-2307  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/353,700  
; FILING DATE: 09-DEC-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: REED, JANET E.  
; REGISTRATION NUMBER: 36,252  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (215) 563-4100  
; TELEFAX: (215) 563-4044  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 10136 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cdna to mRNA  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: HUMAN  
; US-08-353-700-2

Query Match 3.6%; Score 56.2; DB 1; Length 10136;  
Best Local Similarity 62.4%; Pred. No. 1.7e-05;  
Matches 88; Conservative 0; Mismatches 53; Indels 0; Gaps 0;  
Qy 1415 CTGCTGGCAGACAAAGCTATTAAGCTATGGCTTATTACAAATCATTAACCTGTGACA 1474  
Db 9990 CTGTTAGCATTCCTCTACTGCAATGTAATAGTATAAAGCTATGTATATAAAGC 10049  
Qy 1475 CGTAAAGAAATFACTTAGGAATAAAGTAAACGCTCTGAGTAAAAAAGAAAAA 1534  
Db 10050 TTTTGGTAAATGTGTTACAAATTAATAATGCAAGCACTATATAAAAAA 10109  
Qy 1535 AAAAAA 1555  
Db 10110 AAAAAA 10130

RESULT 8  
PCT-US95-16216-2  
; Sequence 2, Application PC/TUS9516216  
; GENERAL INFORMATION:  
; APPLICANT: Yen, Timothy J.  
; APPLICANT: Rattner, Jerome B.  
; TITLE OF INVENTION: Nucleic Acid Encoding a Transiently  
; TITLE OF INVENTION: Expressed Kinetochore Protein, and Methods of Use  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman  
; STREET: 1601 Market Street Suite 720  
; CITY: Philadelphia  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19103-2307





Db 1588 ATAATGTTAAATAAATAATTAATCTGTGAATAAAGATATGATTTTATTCAAACT 1647  
QY 1487 ACTTAGGAATAAAGCTAAACGCTCTGAGTAAATAAAAAAAAAAAAAAAAAAAAA 1546  
Db 1648 TGTCAATATATAAA 1707  
QY 1547 AAAAAAAAA 1555  
Db 1708 AAAAAAAAA 1716

RESULT 13  
US-09-182-816-24/c  
; Sequence 24, Application US/09182816  
; Patent No. 6145342  
; GENERAL INFORMATION:  
; APPLICANT: Wisniewski, Nancy  
; APPLICANT: Silver, Gary M.  
; APPLICANT: Lo, Katherine C.  
; APPLICANT: Brandt, Kevin S.  
; TITLE OF INVENTION: NOVEL FLEA EPOXIDE HYDROLASE NUCLEIC ACID MOLECULES,  
; TITLE OF INVENTION: PROTEINS AND USES THEREOF  
; FILE REFERENCE: FC-3-C1  
; CURRENT APPLICATION NUMBER: US/09/182,816  
; CURRENT FILING DATE: 1998-10-29  
; EARLIER APPLICATION NUMBER: 08/989,510  
; EARLIER FILING DATE: 1997-12-12  
; NUMBER OF SEQ ID NOS: 31  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 24  
; LENGTH: 1736  
; TYPE: DNA  
; ORGANISM: Ctenocephalides felis  
US-09-182-816-24

Query Match 3.6%; Score 55.4; DB 3; Length 1736;  
Best Local Similarity 64.3%; Pred. No. 1.1e-05;  
Matches 83; Conservative 0; Mismatches 46; Indels 0; Gaps 0;  
QY 1427 ACAAGTTATAAGCTATGGCTTATTACAAATCAATTAACCTGTACACGCGTAAGAAAT 1486  
Db 149 ATAATGTTAAATAAATAATGTAATTAATCTGTGAATAAAGATATGATTTTATTCAAACT 90  
QY 1487 ACTTAGGAATAAAGCTAAACGCTCTGAGTAAATAAAAAAAAAAAAAAAAAAAAA 1546  
Db 89 TGTCAATATATAAA 30  
QY 1547 AAAAAAAAA 1555  
Db 29 AAAAAAAAA 21

RESULT 14  
US-09-471-528-22  
; Sequence 22, Application US/09471528  
; Patent No. 6153397  
; GENERAL INFORMATION:  
; APPLICANT: Wisniewski, Nancy  
; APPLICANT: Silver, Gary M.  
; APPLICANT: Lo, Katherine C.  
; APPLICANT: Brandt, Kevin S.  
; TITLE OF INVENTION: FLEA EPOXIDE HYDROLASE PROTEINS AND USES THEREOF  
; FILE REFERENCE: FC-3-C1-1  
; CURRENT APPLICATION NUMBER: US/09/471,528  
; CURRENT FILING DATE: 1999-12-27  
; EARLIER APPLICATION NUMBER: 09/182,816  
; EARLIER FILING DATE: 1998-10-29  
; EARLIER APPLICATION NUMBER: 08/989,510  
; EARLIER FILING DATE: 1997-12-12  
; NUMBER OF SEQ ID NOS: 35  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 22  
; LENGTH: 1736

; TYPE: DNA  
; ORGANISM: Ctenocephalides felis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (159)..(1553)  
US-09-471-528-22

Query Match 3.6%; Score 55.4; DB 3; Length 1736;  
Best Local Similarity 64.3%; Pred. No. 1.1e-05;  
Matches 83; Conservative 0; Mismatches 46; Indels 0; Gaps 0;  
QY 1427 ACAAGTTATAAGCTATGGCTTATTACAAATCAATTAACCTGTACACGCGTAAGAAAT 1486  
Db 1588 ATAATGTTAAATAAATAATGTAATTAATCTGTGAATAAAGATATGATTTTATTCAAACT 1647  
QY 1487 ACTTAGGAATAAAGCTAAACGCTCTGAGTAAATAAAAAAAAAAAAAAAAAAAAA 1546  
Db 1648 TGTCAATATATAAA 1707  
QY 1547 AAAAAAAAA 1555  
Db 1708 AAAAAAAAA 1716

RESULT 15  
US-09-471-528-24/c  
; Sequence 24, Application US/09471528  
; Patent No. 6153397  
; GENERAL INFORMATION:  
; APPLICANT: Wisniewski, Nancy  
; APPLICANT: Silver, Gary M.  
; APPLICANT: Lo, Katherine C.  
; APPLICANT: Brandt, Kevin S.  
; TITLE OF INVENTION: FLEA EPOXIDE HYDROLASE PROTEINS AND USES THEREOF  
; FILE REFERENCE: FC-3-C1-1  
; CURRENT APPLICATION NUMBER: US/09/471,528  
; CURRENT FILING DATE: 1999-12-27  
; EARLIER APPLICATION NUMBER: 09/182,816  
; EARLIER FILING DATE: 1998-10-29  
; EARLIER APPLICATION NUMBER: 08/989,510  
; EARLIER FILING DATE: 1997-12-12  
; NUMBER OF SEQ ID NOS: 35  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 24  
; LENGTH: 1736  
; TYPE: DNA  
; ORGANISM: Ctenocephalides felis  
US-09-471-528-24

Query Match 3.6%; Score 55.4; DB 3; Length 1736;  
Best Local Similarity 64.3%; Pred. No. 1.1e-05;  
Matches 83; Conservative 0; Mismatches 46; Indels 0; Gaps 0;  
QY 1427 ACAAGTTATAAGCTATGGCTTATTACAAATCAATTAACCTGTACACGCGTAAGAAAT 1486  
Db 149 ATAATGTTAAATAAATAATGTAATTAATCTGTGAATAAAGATATGATTTTATTCAAACT 90  
QY 1487 ACTTAGGAATAAAGCTAAACGCTCTGAGTAAATAAAAAAAAAAAAAAAAAAAAA 1546  
Db 89 TGTCAATATATAAA 30  
QY 1547 AAAAAAAAA 1555  
Db 29 AAAAAAAAA 21

Search completed: June 2, 2003, 04:56:39  
Job time : 104 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 2, 2003, 04:13:34 ; Search time 842 Seconds  
(without alignments)  
2490.152 Million cell updates/sec

Title: US-09-830-980-2  
Perfect score: 1555  
Sequence: 1 aattcccaaaatgcagga.....aaaaaaaaaaaaaaaaaaaa 1555

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 845702 seqs, 674182571 residues

Total number of hits satisfying chosen parameters: 1691404

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications NA.:

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq.\*
- 11: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*
- 12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq.\*
- 13: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*
- 14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	478	30.7	1859	9 US-10-132-744A-5	Sequence 5, Appli
2	466.2	30.0	1928	9 US-10-132-744A-1	Sequence 1, Appli
3	249.4	16.0	1170	9 US-10-132-744A-3	Sequence 3, Appli
4	155.8	10.0	1722	9 US-09-774-639-34	Sequence 34, Appl
5	155.8	10.0	1722	9 US-09-969-730-87	Sequence 87, Appl
6	133.4	8.6	418	9 US-09-918-995-11539	Sequence 11539, A
7	120	7.7	445	9 US-09-918-995-37909	Sequence 37909, A
8	79.4	5.1	3025	10 US-09-764-853-29	Sequence 29, Appl
9	59.6	3.8	2431	10 US-09-794-257-4	Sequence 4, Appli
10	59.4	3.8	2431	10 US-09-925-300-464	Sequence 464, App
11	59	3.8	2674	9 US-10-003-295-1	Sequence 1, Appli
12	58.4	3.8	453	9 US-09-918-995-30837	Sequence 30837, A
13	57.6	3.7	1602	9 US-09-997-279-19	Sequence 19, Appl
14	57.6	3.7	1602	10 US-09-813-358-19	Sequence 19, Appl
15	57.4	3.7	393	10 US-09-960-352-4582	Sequence 4582, Ap
16	57.4	3.7	1221	9 US-10-237-381-2	Sequence 2, Appli
17	57.4	3.7	1241	9 US-09-822-846-582	Sequence 582, App
18	57	3.7	266	9 US-10-091-483-42	Sequence 42, Appl
19	57	3.7	266	10 US-09-764-846-42	Sequence 42, Appl

20.	56.8	3.7	1378	9 US-10-098-841-121	Sequence 121, App
21	56.8	3.7	1522	9 US-10-098-841-122	Sequence 122, App
22	56.8	3.7	3466	9 US-10-042-141-35	Sequence 35, Appl
23	56.8	3.7	3466	10 US-09-726-643-35	Sequence 35, Appl
24	56.4	3.6	447	9 US-09-918-995-13767	Sequence 13767, A
25	56.4	3.6	475	9 US-09-918-995-31138	Sequence 31138, A
26	56.4	3.6	1133	9 US-10-282-048-1	Sequence 1, Appli
27	56.4	3.6	3449	9 US-09-925-299-225	Sequence 225, App
28	56.4	3.6	3449	10 US-09-925-299-225	Sequence 225, App
29	56.2	3.6	347	10 US-09-960-352-1096	Sequence 1096, Ap
30	56.2	3.6	2732	9 US-09-759-056-1	Sequence 1, Appli
31	56.2	3.6	2732	9 US-09-901-812-1	Sequence 1, Appli
32	56.2	3.6	2732	9 US-10-227-884-79	Sequence 79, Appl
33	56.2	3.6	2732	9 US-10-230-163-79	Sequence 79, Appl
34	56.2	3.6	2732	9 US-10-218-631-79	Sequence 79, Appl
35	56.2	3.6	2732	9 US-10-230-338-79	Sequence 79, Appl
36	56.2	3.6	2732	9 US-10-230-414-79	Sequence 79, Appl
37	56.2	3.6	2732	9 US-10-216-159A-79	Sequence 79, Appl
38	56.2	3.6	2732	9 US-10-218-849-79	Sequence 79, Appl
39	56.2	3.6	2732	9 US-10-227-873-79	Sequence 79, Appl
40	56.2	3.6	2732	9 US-10-227-883-79	Sequence 79, Appl
41	56.2	3.6	2732	9 US-10-219-076-79	Sequence 79, Appl
42	56.2	3.6	2732	9 US-10-230-434-79	Sequence 79, Appl
43	56.2	3.6	2732	9 US-10-219-003-79	Sequence 79, Appl
44	56.2	3.6	2732	9 US-10-219-075-79	Sequence 79, Appl
45	56.2	3.6	2732	9 US-10-219-464-79	Sequence 79, Appl

ALIGNMENTS

RESULT 1

US-10-132-744A-5

; Sequence 5, Application US/10132744A

; Publication No. US20030027261A1

; GENERAL INFORMATION:

; APPLICANT: Utku, Nalan

; TITLE OF INVENTION: No. US20030027261alel genes Tzap7/A, Tzap7/B and Tzap7 involve

; FILE REFERENCE: Utku-4 CON

; CURRENT APPLICATION NUMBER: US/10/132.744A

; CURRENT FILING DATE: 2002-04-24

; PRIOR APPLICATION NUMBER: PCT/EP00/10670

; PRIOR FILING DATE: 2000-10-30

; PRIOR APPLICATION NUMBER: 60/185,016

; PRIOR FILING DATE: 2000-02-25

; PRIOR APPLICATION NUMBER: 60/162,675

; PRIOR FILING DATE: 1999-11-01

; NUMBER OF SEQ ID NOS: 6

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 5

; LENGTH: 1859

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: miscellaneous feature

; LOCATION: 1154; 379

; OTHER INFORMATION: ATCG; variable amino acid

US-10-132-744A-5

Query Match 30.7%; Score 478; DB 9; Length 1859;

Best Local Similarity 59.8%; Pred. No. 3.4e-118;

Matches 837; Conservative 0; Mismatches 556; Indels 6; Gaps 2;

QY 54 GATCAGCGCGCGCTCGTTTACGGGGGAGGAGCGCGCCCAATGACCTGACCTGCCGC 113

Db 70 GCTAGTCAGTTTCCAGGATGAGGCGCGGACGCTGCTGGGTTCCTCCCTTCGAGTCCCGT 129

QY 114 AGGAATCACTACCAGCAATTTGGACTGATTTCGAACGCGCTGCTGAAACAGGAGGAGC 173

Db 130 GGACATACCCCGGACAGGCTGACGCTGTGTGCAACGCGCTACTGGCCAGGAGGATCC 189

QY 174 CACTCCATATTTGTTTTTCGTGGCGGAGGATGAGATCAAGAGAGCTGGAGGACACGTT 233

```
Db 190 CTTGCCACTGGCTTTCTTTGTTCCACGATGCTGAGATCTCTCCTCACTGGGGAACAGCTT 1249
QY 234 GGACTTGGCGTCACTGAGACACACGAAACGTCATCGATATTTGTATATCACTCCACACAGCGGT 293
Db 250 GGAGTCCACGCGAGTGGACAGAGAGGTCTTAGACATCATCTACACGCCACAGGCTAT 309
QY 294 TTTCAAAGTGGCCAGTGCACAAAGATGCACGAGTTCATCCCGGGACACCGCCGAGGCTGT 353
Db 310 CTTTCAGAGTCCGGGCTGTGACTCGCTGCACCAAGCTCTTGGAGGCTCAAGTGGAGGCACT 369
QY 354 GGTTCGTGTAATTTTCAGCCCGATGTCATCTCCGACAGTGGAGTGGCGACACAC 413
Db 370 CATTTCTGTGCTTCAGCCCTACGGGAAAGTACTCTGGCCAGTGGCTCTGGAGAGACCC 429
QY 414 AGTCGGATTGGGATCTTAACACAGACACACCGACTTCACCTGCACAGTTCATAAGCA 473
Db 430 CTTGCGCTTCGGATCTCAGACAGACACACACATTTTCATATGCAAGGGACACAGACA 489
QY 474 GTGGGTTCTGTGGTATCTGGGCTCCGGATGGCAACGGTTGGCCAGCGTTCAAGC 533
Db 490 CTGGGTCCTTAGTATATCTGTCTCCAGATGGCAAGAGCTGGCTCAGGCTGCAAGAA 549
QY 534 GGGCTCTAATCATCTGGGACCGGAGCGGTCAAGAGAGAGGGGCGACCTTGAGTGG 593
Db 550 TGGCCAGATTCTCTCTGGGACCAAGCACAGGGAAGCAGGTGGGCGAGGACCTCGTGG 609
QY 594 GCACAAGAAACATCACTGCTGCTGGGAAACGATATCATCGGATCCGGAGTGCAG 653
Db 610 CCACAGCAAGTGTACACAGGCTGAGCTGGGAGCCCTCCATCGGAACCTTGAGTCCG 669
QY 654 GAAACTTGTCTCGGCTGGAGCGGGGACTCGCGGATTTGGGAGTAAATTTGGGCCA 713
Db 670 CTATGTGGCAGCAGCTCCAGGATGGGAGTGGCGGATCTGGGACACAATTCGAGCGG 729
QY 714 GTGCTTATGAATTCGCGGACACACAATGTGTGACAGCAGTGTGAGTGGGTGGAGC 773
Db 730 CTTGTGAGCGCATCTCACCGGGCACACCGAGTGGTCACTGTCTCCGGTGGGAGGGA 789
QY 774 GGGCTTATTTATACATCTCAAGATGCACAGTGAAGATGTGGCAGCAGCTGATGG 833
Db 790 CGGCTCTCTACTCTGCTCCAGGACCGCACCATCAAGTCTGGAGAGCTCATGACGG 849
QY 834 AATCTGTGCGGAGCTTCTTGCCGAAGTCTACTGGGTAACAACATGTGCTGTAGCAG 893
Db 850 TGTGCTGTGCGGAGCTGTGAAGGCCACCGCCACTGGGTGAACACCACTGCGCTCAGC 909
QY 894 CGAATACGCTCTGCGCACTGCTTCCATTCCTGCGGTGAAGGATCGCTCCAAGAGCC---A 950
Db 910 TGACTATGCTGCGCACTGGGCGCTTTGAACCTGCTGAGGCGCTCAGTTAATCCCAAGA 969
QY 951 CCTCACTTTGACACTGAGGAATTCAGGAATCTGCTTGAAGCGCTACACAGCGCGTGTG 1010
Db 970 CTTCAAGATCCTTGCAGGAGTTGAAGGAGAGGGCTCTGAGCGGATACAACTCTGTCG 1029
QY 1011 CCTGACGAGGTGGAGTCTGTTCTGTTGCGATGACACACCTCTATCTGTGG-- 1068
Db 1030 GGGCCAGGTCCTGACAGAGGCTGTGCTGCTGCGGACGACTTCACTTCTGCTGTGTC 1089
QY 1069 -CGGAACAACAGACAGAGTGCCTGTGAGCGCATGACAGGCGACACAGACGTGTCAACGA 1127
Db 1090 CCACAGCAGAGCAAAAGCCTCTCACTCGGATGACAGACACCAAGCTCTCATCAACCA 1149
QY 1128 TGTGAATATTCGGCGGATGTAAGCTAATTCGCTCTCTTCAATTTGACAAGTCACTGG 1187
Db 1150 GTGTGTTCTCTCTGACTCCCGCATCGTGGCTAGTGCCTCTTTCACAAGTCCATCAA 1209
QY 1188 TCTGTGGCGAGCAGCATGTCAGTACATGCGCACTTCGCGGCTCATGTGCGAGCTGT 1247
Db 1210 GCTGTGGGATGGCAGGACGGCAGTACTGCTTCCCTACCGCGGCGACCTGCTGCTG 1269
QY 1248 TTACAGGTTGCTGTGTCGCGGACTCCCGCTTGTATGTTTCCGGCAGCAAGACTCAAC 1307
```

```
Db 1270 GTACCAGATTGGTGGTCAGCTGACAGTCCGTCTCTGTGTGAGCGGACAGTGTACAGCAC 1329
QY 1308 TCTAAAGATATGAGTGTGTCAGACGAAGAACTGGCAGAGAGTGCCTGTGACATGCGGA 1367
Db 1330 ACTGAAGTGTGGATGTGAAGGCCCAAGAGCTTGGCATATGACCTGCCGCGCACGCGGA 1389
QY 1368 TGAGGTGTTCCGAGTGGAGTGGCGCGCGCATGCTGTAGACTTGCCTCTGTTGGCAAGA 1427
Db 1390 TGAGGTATATGCTGTGACTGGAGTCCAGATGGCCAGAGTGGCAAGTGTGGGAAGGA 1449
QY 1428 CAAAGTATATAAGCTATGG 1446
Db 1450 CAAATGCTCCGGATATGG 1468

RESULT 2
US-10-132-744A-1
; Sequence 1, Application US/10132744A
; Publication No. US20030027261A1
; GENERAL INFORMATION:
; APPLICANT: Utku, Nalan
; TITLE OF INVENTION: No. US20030027261A1 genes Tzap7/A, Tzap7/B and Tzap7 involved
; FILE REFERENCE: Utku-4 CON
; CURRENT APPLICATION NUMBER: US/10/132,744A
; CURRENT FILING DATE: 2002-04-24
; PRIOR APPLICATION NUMBER: PCT/EP00/10670
; PRIOR FILING DATE: 2000-10-30
; PRIOR APPLICATION NUMBER: 60/185,016
; PRIOR FILING DATE: 2000-02-25
; PRIOR APPLICATION NUMBER: 60/162,675
; PRIOR FILING DATE: 1999-11-01
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1928
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (92)..(1543)
US-10-132-744A-1

Query Match 30.0%; Score 466.2; DB 9; Length 1928;
Best Local Similarity 59.3%; Pred. No. 5,1e-115;
Matches 830; Conservative 0; Mismatches 563; Indels 6; Gaps 2;

QY 54 GATACAGCGCGCTCGTTTACAGCGGAGAGAGCGCGCCCAATCGACCTGCCGCG 113
Db 139 GCTAGTGGAGTTCAGAGTAGGGCGGCGAGCTGCTGGGTTCCTCCGCTTCGACGTGCCGT 198
QY 114 AGGAATCACTACCCAGCAATTTGGGACTGATTTGCAACGCGCTGCTGAAAACGAGGAAGC 173
Db 199 GGACATCACCCCGGACAGGCTGCAGCTCGTTTGAACGCGCTACTGCCCCAGAGATCC 258
QY 174 CACTCCATATTTTCTTTTCTGGGCGAGATGACATCAAGAGAGCTCGGAGGACACGTT 233
Db 259 CCTGTACTGGCTTCTTTTCTCCAGATGCTGAGATGTTTCTCTACTCGGAGAGACGTT 318
QY 234 GGACTTGGCTCAGTGGACACCCGAAAGCTGATCGATTTGTGTATTCACCCACAGCGCGT 293
Db 319 GGAGTCCCGGACAGTGGAGACAGAGAGTCTTAGACATCATTTTACCACACAGAGCTGT 378
QY 294 TTTCAAAGTCCGCCAGTGCAGAGATGCAGAGTCCATGCCGGGACAGCGCGAGGCTGT 353
Db 379 GTTCAAGTTCGTGCTGTACACAGATGACAGTCTATTGGAGGAGACACCGAGGCTGT 438
QY 354 GGTTCGCTGAATTTACGCCCGGATGTTGCTCTCTCGCCAGTGGAGTGGCGACACAC 413
Db 439 TATTAGTAGAGCTTCAGCCCCAAGTATTTGGCAAGTATTTGGCAAGTGTCTTGGGAGACTAC 498
QY 414 AGTCGATTTGGGATCTTAACACAGAGACAGCGCACTTCACCTGCACAGTGTCAAGCA 473
```



```
Db 499 AGTCGCTTTTGGGATCTCAGCACAGAACTCCACATTTTACATCTAAAGGGCATACACA 558
QY 474 GTGGGTTCTGTGGGTATCTCTGGGCTCCGGATGGCAAAACGGTTGGCCAGCGGTGCAAAAGC 533
Db 559 CTGGGTTCTAGTATGCTTGGTCTCCAGATGGCAAAACCTTGCCTCAGGATGTAAAAA 618
QY 534 GGGCTCTATAATCATCTGGGACCGGAGAGCGGTGCAGCAAGGGCGGACCCCTTGAGTGG 593
Db 619 TAGTCAGATCTTCATTTGGGACCCAGCAGGAGGAGGATGGCAAAACCATTAACAG 678
QY 594 GCACAGAAACACATCACTGCTCGCTGGGAAACGGTATCATCGCGATCCGGAGTGCAG 653
Db 679 GCACCTAAAGTGGATTACATGGGTGTGGGAAACCTCTCCACCTGAACCCAGAGAGCGG 738
QY 654 GAACTTGCTTCCGCCAGTGGAGACGGGACTGCGGGATTGGGACACATAAAATGGGCCA 713
Db 739 ATACCTAGCCAGTGCCTCCAGCGCGCGTCCAGCCGATCTGGGACACAACTCGAGCGG 798
QY 714 GTGCCCTTATGAACATTCGCGGACACACAATGCTGTGACAGCAGTGAGTGGGTGAGC 773
Db 799 CTGTGAGCCATCTCACCAGGACACACAGTCGGTACCTGTCTCCGGTGGGAGGGA 858
QY 774 GGGCTTATTTATATCTCTCCAAAGATCCGACAGTGAAGATGTGGCGAGCAGCTGATGG 833
Db 859 CGGGCTTCTACTCTGCTCCAGGACCGCACCATCAAAAGTCTGGAGAGCTCATGCGG 918
QY 834 AATCTTGTCGGGAGTCTCTGCGCAAGCTCAGTGGGTAAACACATTCGCTGAGCAC 893
Db 919 TGTGCTGTGCGGACTCTGACAGGCCACGGCCACTGGGTGAACACATGACCTTCAGCAC 978
QY 894 CGATTACGCTTGGGCACTGCTCCATTCATCCGTTGGAAGGATCGCTCCAAAGAGCC---A 950
Db 979 TGACTATGCTTGGCAGCTGGGCGCTTTGAACCTGTGTGAGGCGCTCAGTTAATCCCAAGA 1038
QY 951 CCTCAGTTTGAGCACTGAGGAATTCGAGGAATCTGCCTTGAAGCGCTACCAGGCGCTGTG 1010
Db 1039 CTTCCAGGATCTTTCAGGAGTTGAAGGAGAGGCTCTGAGCGATACAACTCTGCTGGC 1098
QY 1011 CCTGACGAGTGGAGTGGTGTCTCTGTTTGGATGACAAACCCCTCTATCTGTGG--- 1068
Db 1099 GGGCCAGGCTCCAGAGAGGCTGGTGTCTGGCTCCGACGACTTCACCTTATTCCTGTGGTC 1158
QY 1069 -CGGAACACCAAGCAAGTGGTGGGCGCATGACAGGCGCACCAACCTGTGTCAACGA 1127
Db 1159 CCAGCAGAGACAAAGAGCTCTCACTCGGATGACAGGACACCAAGCTCTCATCAACCA 1218
QY 1128 TGTGAATATTCGCGGATGTAAAGCTAATTCGCTCTGCTTCATTTGACAAGTCACTGCG 1187
Db 1219 GGTGCTCTTCTCTCTGACTCCCGCATCTGGCTAGTGTGCTCTCTTGACAAGTCCATCAA 1278
QY 1188 TCTGTGGCGAGCCAGCGATGTGATACATGGCCACCTTCCGGGGTTCATGTGAGGCTGT 1247
Db 1279 GCTGTGGATGGCAGGAGGCAAGTACCTGCTTCCCTACGCGGCCACCTGGCTGCCGT 1338
QY 1248 TTAACGCTTGGCTGCGGAGCTCCCGCTGCTGATTTGTTCCGGCAGCAAGACTCAAC 1307
Db 1339 GTACAGATTTGGTGTGAGCTGACAGTGGCTTCTGCTGAGCGGCGAGCAGTACAGCAC 1398
QY 1308 TCTAAAGTATGAGTGTGACAGCAAGAACTGGCAGGAGTGTGCTGGATGCGGA 1367
Db 1399 ACTGAAGTGTGGGATGTGAAGGCCAGAGCTGGCCATGGACCTGCCGCGCACCGGA 1458
QY 1368 TGAGGTTCGAGTGGACTGGGGGCGGATGGCTCTAGAGTTCCTCTGTTGGGCAAGA 1427
Db 1459 TGAGGTATATGCTGTTGACTGGAGTCCAGATGGCCAGAGAGTGGCAAGTGTGGGAAGA 1518
QY 1428 CAAAGTTATAAGCTATG 1446
Db 1519 CAAATGCTCCGATATGG 1537
```

RESULT 3

US-10-132-744A-3

```
; Sequence 3, Application US/10132744A
; Publication No. US20030027261A1
; GENERAL INFORMATION:
; APPLICANT: Utlu, Nalan
; TITLE OF INVENTION: No. US20030027261A1el genes Tzap7/A, Tzap7/B and Tzap7 involve
; FILE OF INVENTION: activation and uses thereof
; FILE REFERENCE: Utlu-4 CON
; CURRENT APPLICATION NUMBER: US/10/132,744A
; PRIOR FILING DATE: 2002-04-24
; PRIOR APPLICATION NUMBER: PCT/EP00/10670
; PRIOR FILING DATE: 2000-10-30
; PRIOR APPLICATION NUMBER: 60/185,016
; PRIOR FILING DATE: 2000-02-25
; PRIOR APPLICATION NUMBER: 60/162,675
; PRIOR FILING DATE: 1999-11-01
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 1170
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3)..(785)
US-10-132-744A-3
```

```
Query Match 16.0%; Score 249.4; DB 9; Length 1170;
Best Local Similarity 59.3%; Pred. No. 7.6e-57;
Matches 462; Conservative 0; Mismatches 311; Indels 6; Gaps 2;
```

```
QY 674 GAGACGGGAGTCCCGGATTTGGGACGTAAATTTGGCCAGTCCCTTATGAACATGCGG 733
Db 1 GCGGCGCGCTGACCGCGGCTGTGGACACAACTGTGAGCGCGCTGTGAGCGCATCTCCACG 60
QY 734 GACACAAATCTGTGACAGCAGTGCAGTGGGTGGGAGCGGCGCTTATTATACATCT 793
Db 61 GGCACACCCAGTCTGGTCTGCTCCGGTGGGAGGGGACGGGCTTCTTACTCTGCT 120
QY 794 CCAAGATCGCACAGTGAAGATGTGGCAGCAGTGTGAATCTTTGTGCCGAGCTTCT 853
Db 121 CCCAGGACCCGACCATCAAGTCTGGAGAGCTCATGACGGTGTGCTGTGCTGCCGACTCG 180
QY 854 CTGGCAAGCTCACTGGGTAAACAACTTGGCTGAGCAGCCGATTAAGTCTGCTGCGCATG 913
Db 181 AAGGCGAGGCGGCTGGGTGAACACCATGCGCTCAGCATGAGTACTGATGCGCTGCCACTG 240
QY 914 GTCCATTCATCCGGTGAAGGATCGCTCCAAAGACC---ACCTCAGTTTGGACACTGAGG 970
Db 241 GGGGCTTTGAACCTGTCTGAGGCGCTCAGTTAATCCCAAGACCTCCAAGGATCTTTCAGG 300
QY 971 AATTGCAAGGATCTGCTTGAAGCGCTACCGGCGGTGTGCTGACGAGGTGGAGTCCG 1030
Db 301 AGTTGAAGGAGAGGGCTCTGAGCGGATACACCTCGTGGCGGGCCAGGGTCCAGAGAGGC 360
QY 1031 TGGTTTCTGTTCGGATGACACACCTCTATCTGTGG---CGGAACACCAAGACAAAGT 1087
Db 361 TGGTGTCTGGCTCCGAGACTTCACCTTATTCCTGTGGTCCCGCAGAGGACAAAAGC 420
QY 1088 GCGTTGAGCGCATGACAGGCGCACAGAACGTTGGTCAACGATGTGAAATATTTCGCCGGATG 1147
Db 421 CTCCTCCTCGGATGACAGGACACCAAGCTCTCATCAACAGGTGCTTCTCTCTGCTGACT 480
QY 1148 TAAAGCTAATTGGTCTGCTTTCATTTGACAAGTCACTGCGTCTGTGGCAGGACCGGATG 1207
Db 481 CCCGATCGTGGGTAGTGTCTCTTTGACAAGTCCATCAAGCTGTGGGATGGCAGGAGCG 540
QY 1208 GTCAATACATGGCCACCTCCCGGGGTGATGTGACAGGTGTTTACAGGTTGCTGCTCGCG 1267
Db 541 GCAAGTACCTGGCTTCCCTACGCGGCCAGCTGGCTGCGGTGTACCAGATTGCTGTGTCAG 600
QY 1268 CGGACTCCCGCTGTGATTTGTTCCGGCAGCAAGACTCACTCTAAAGATATCGGAGTGTGC 1327
Db 601 CTGACAGTCCGCTCTCTGGTGGTGGCAGCGCAGCTGACAGCACACTGAAAGGTGTGGGATGTA 660
```



```
; PRIOR APPLICATION NUMBER: 60/055,386
; PRIOR FILING DATE: 1997-08-05
; NUMBER OF SEQ ID NOS: 373
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 87
; LENGTH: 1722
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: SITE
; LOCATION: (2)
; OTHER INFORMATION: n equals a.t.g, or c
; NAME/KEY: SITE
; LOCATION: (413)
; OTHER INFORMATION: n equals a.t.g, or c
US-09-969-730-87

Query Match
Best Local Similarity 10.0%; Score 155.8; DB 9; Length 1722;
Matches 326; Conservative 1; Mismatches 224; Indels 7; Gaps 3;

QY 896 ATTACGTCCTCGGCACATGTCATTCATTCGAGGATCGCTCCAAAGACC---ACC 952
Db 3 ACTATGCCCTCGGCACATGTCATTCGAGGCTCTGAGGCTCAGTTAATCCCAAGACC 62

QY 953 TCAGTTGAGCACTGAGGAATTCAGGATTCGCTTGAAGCCCTACCGGCGGTGCC 1012
Db 63 TCCAAAGGATCTTCAGGAGTTGAAGGAGAGGGCTCTGAGCCGATACACCTCGTGGG 122

QY 1013 CTGAGGAGTGGAGTCTGCTGTTTCTCGGATGACACACCTCTATCTGTGG---C 1069
Db 123 GCCAGGTCACAGAGAGGCTGTGTCTGCTCGGAGGCTTCACTTCTGCTGCTGCC 182

QY 1070 GGAACACCAACCAAGTCGTTGAGCGCATGACAGGCGCAGACAGTGTCAAGATG 1129
Db 183 CAGCAGAGGACAAAGCCCTCTCACTCGGATGACAGGACCAAGCTCTCATCAACGAG 242

QY 1130 TGAATATTCGCGGATGTAAGACTAATTCGCTGCTGCTTCAATTTGACAACTAGTCGTC 1189
Db 243 TGTCTTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 302

QY 1190 TGTGGGAGCAGGATGGTCAGTACATGCGCAGCTTCCGGGCTCATGTGCGAGCTGTTT 1249
Db 303 TGTGGGATGCGAGGAGGCGCAAGTACCTGCTTCCCTACGCGGCCAGCTGGCTGGTGT 362

QY 1250 ACAGGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1308
Db 363 ACCAGATTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 422

QY 1309 CTAAGATGATGAGTGTGACAGCAAGAACTGGCAGAGGAGCTGCTGCTGCTGCTGCTGCT 1368
Db 423 CTGAAGTGTGGATGTGAAGGCCCAAGAGCTGGCCATGGACCTGCCCGGCCAGCGGAT 482

QY 1369 GAGGTTCCGAGTGGAGTGGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1428
Db 483 GAGGTTATGCTGTTGACTGGAGTCCAGATGGCCAGAGTGGCAAGTGGTGGGAAGGAC 542

QY 1429 AAAGTTATAAAGCTATGG 1446
Db 543 AAATGCTTCGGATATGG 560

RESULT 6
US-09-918-995-11539
; Sequence 11539, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 37909
; LENGTH: 445
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-918-995-37909

Query Match
Best Local Similarity 7.7%; Score 120; DB 9; Length 445;
Matches 254; Conservative 0; Mismatches 180; Indels 6; Gaps 2;

QY 761 GATGGGTGGAGCGGCTTTATTATATCATCTTATATATCTTCAAAAGATCGCACAGTGTGGC 820
Db 3 GGTGGGAGGAGCGGCTTCTCTACTCTGCTTCCCGAGGACCGCACCATCAAGTCTGGA 62

QY 821 GAGCAGCTGATGGAATCTTCTGCGGACAGCTTCTCTGCGCAAGCTCTACTGGGTAACAACA 880
```

```
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 11539
; LENGTH: 418
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-918-995-11539

Query Match
Best Local Similarity 8.6%; Score 133.4; DB 9; Length 418;
Matches 236; Conservative 0; Mismatches 171; Indels 0; Gaps 0;

QY 528 CAAAGCGGCTCTATAATCATCTGGGACCGGAGCGGTGACGAGAGGGCGGACCCCTT 587
Db 1 CAAGAATGGCCAGATTCCTCTGGGACCAAGCAACAGGAGGCGGTGGGCTTTACCCCT 60

QY 588 GAGTGGGCAACAAGACATCAACTGCTGCTGGGAAACGATATCATCGCGATCCGGA 647
Db 61 CGTGGCCACAGCAAGTGTGATCAGAGCTGAGCTGGAGCCCTCCATGCGAACCCTGA 120

QY 648 GTCCAGGAACTTGTCTTCCGCGAGTGGAGAGCGGGAGTGGCGGATTTGGAGCTAAATTT 707
Db 121 GTGCCGCTATGTGGCGAGCAGCTCCAAGGATGGCAGCGTGGGATCTGGGACACACTGC 180

QY 708 GGGCAGTGCCTTATGAACATTTGCGGACACACAAATGCTGTGACAGCAGTGAATGGG 767
Db 181 AGGCCGCTGTGAGCGCATCTCACCGGCACACCCAGTCGGGACCTGCTCCGTTGGG 240

QY 768 TGGAGCGGCGCTTATTTATATCTTCCAAAGATCGGACAGTGAAGATGTGGCAGCAGC 827
Db 241 AGGGAGCGGCTTCTCTAGCTTGTCCAGGACCGGACCATCAAAAGTCTGGATAGCTCA 300

QY 828 TGATGGAATCTTGTGCGGAGCTTCTCTGCGCAAGCTCTACTGGGTAACAACATTTGCCCT 887
Db 301 TGACGCTGTCTGTGCGGAGCTTCTTGACGGCCAGCGGCACTGTGTGAACACCATGGCCT 360

QY 888 GAGCAGCATTTAGCTCTGCGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 934
Db 361 CAGCAGTACTGCTGCGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 407

RESULT 7
US-09-918-995-37909
; Sequence 37909, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 37909
; LENGTH: 445
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-918-995-37909

Query Match
Best Local Similarity 7.7%; Score 120; DB 9; Length 445;
Matches 254; Conservative 0; Mismatches 180; Indels 6; Gaps 2;

QY 761 GATGGGTGGAGCGGCTTTATTATATCATCTTATATATCTTCAAAAGATCGCACAGTGTGGC 820
Db 3 GGTGGGAGGAGCGGCTTCTCTACTCTGCTTCCCGAGGACCGCACCATCAAGTCTGGA 62

QY 821 GAGCAGCTGATGGAATCTTCTGCGGACAGCTTCTCTGCGCAAGCTCTACTGGGTAACAACA 880
```

Db 63 GAGCTCATGACGGTGTCTGTCGGGACTCTGCAAGGCCACCGCCACTGGGTGAACACCA 122  
QY 881 TTGGCCTGAGCAGCCGATACGTCCTGCGCAGCTATGTCATTCATCCATCCGGTGAAGGATCGCT 940  
Db 123 TGGCCCTCAGCAGTACTATGCCCCTGCGCACTGGGGCTTTGAACCTGCTGAGGCGCTCAG 182  
QY 941 CCAAGGCC---ACCTCAGTTTGAGCACTGAGGAATTCAGAGGAATCTGCTTGAAGCGCT 997  
Db 183 TTAATCCCAAGACCTCCAAAGGATCTTCAGAGAGTTGAAGGAGAGGGCTCTGAGCCGAT 242  
QY 998 ACCAGGCGGTGCGCTGAGAGGTGGAGTGCCTGCTGCTTCTGTCGATGACAAACCC 1057  
Db 243 ACAACCTCTGCGGGCCAGGTGTCAGAGAGGCTGGTGTCTGGCTCCGACGACTTCACT 302  
QY 1058 TCTATCTGTGG---CGGAACAACCAAGAACAGTCCGTTGAGCGCATGACAGGGCACCAGA 1114  
Db 303 TATTCCTGTGGTCCCGCAGAGGACAAAAGCCTCTCACTCGGATGACAGGACCAAG 362  
QY 1115 ACGTGGTCAACGATGTGAATATTCGCGGATGTAAAGCTAAATGCTGCTCATTTG 1174  
Db 363 CTCTCATCAACCAAGGTGCTTCTCTCTGACTCCCGCATCGGTAGTGGCTGCTCTTTG 422  
QY 1175 ACAAGTCAGTGCCTGCTGTGG 1194  
Db 423 ACAAGTCCCAAGGTGTGG 442

## RESULT 8

US-09-764-853-29  
; Sequence 29, Application US/09764853  
; Patent No. US20020090672A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: P206  
; CURRENT APPLICATION NUMBER: US/09764,853  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 939  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 29  
; LENGTH: 3025  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (3011)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: SITE  
; LOCATION: (3014)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: SITE  
; LOCATION: (3015)  
; OTHER INFORMATION: n equals a,t,g, or c  
US-09-764-853-29

Query Match 5.1%; Score 79.4; DB 10; Length 3025;  
Best Local Similarity 49.2%; Pred. No. 6.5e-11;  
Matches 209; Conservative 0; Mismatches 216; Indels 0; Gaps 0;  
QY 1025 AGTCGCTGTTTCTCTGTTGGATGACACACCCCTCTATCTGTGGCGGAACAACGAGACA 1084  
Db 222 AGCACTTGCTACTGCTTCTTGGGATACCTTCTCATGCTATGGAATTTCAAGCCACATG 281  
QY 1085 AGTGCCTTGAGCGGATGACAGGCGCAGACGTTGGTCAACGATGTGAATATTCGCGG 1144  
Db 282 CTAGAGCTTACAGATAGTGGTCAACAGGATGTGTAAACAGCGTGCATGTTTCTCCAC 341  
QY 1145 ATCTAAAGCTAATGCGTCTGCTTCAATTTGACAAGTCAGTGGCTGTGTGGCGAGCAGG 1204  
Db 342 ATGGAACCTTATTGGCGGTCTGCCTCCAGAGAGAACCGGTGAGACTCTGGATTCCTGATA 401

QY 1205 ATGCTCAGTACATGAGCCACCTTTCGGGGTCAATGTCAGAGCTGTTTACACGGTTGCCGTGGT 1264  
Db 402 AGAGAGAAAATTTCTCAGAATTTAAAGCTCATACAGCTCCAGTTCGAAGTGTAGACTTTT 461  
QY 1265 CCGCGACTCCCGCTTGATTGTTTCCGGCAGCAAGAGACTCAACTCTAAAAGTATGGAGTG 1324  
Db 462 CAGCTGATGGCCAGTTTCTAGTACAGCTTCTGAAGACAATATCCATAAAGTATGGAGCA 521  
QY 1325 TGCAGACGAGAAACTGGCAGAGAGTGCCTGGACATGCGGATGAGGTGTTCCGGAGTGG 1384  
Db 522 TGTATCGCAGCGCTTCTGTTATTCCTTGTATCGACATACACACTGGGTACGCTGTGCCA 581  
QY 1385 ACTGGCGCGCGATGGCTCTAGAGTTGGCTCTGGTGGCAAGACAAAAGTTATAAAGCTAT 1444  
Db 582 AATTTTCAACCGAGGAAGACTAATTGTGTCATGTAGTGAGGATAAACTATTAAATTT 641  
QY 1445 GGGCT 1449  
Db 642 GGGAT 646

## RESULT 9

US-09-794-257-4  
; Sequence 4, Application US/09794257  
; Patent No. US20020009804A1  
; GENERAL INFORMATION:  
; APPLICANT: Meyers, Rachel  
; TITLE OF INVENTION: 32705, 23224, 27423, 32700, 32712, No. US20020009804A1e1  
; FILE REFERENCE: Human G-Proteins  
; FILE REFERENCE: 35800/209285  
; CURRENT APPLICATION NUMBER: US/09794,257  
; CURRENT FILING DATE: 2001-02-27  
; PRIOR APPLICATION NUMBER: 60/185,606  
; PRIOR FILING DATE: 2000-02-29  
; NUMBER OF SEQ ID NOS: 16  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 4  
; LENGTH: 1023  
; TYPE: DNA  
; ORGANISM: homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (245)...(886)  
; NAME/KEY: misc.feature  
; LOCATION: (1)...(1023)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-794-257-4

Query Match 3.8%; Score 59.6; DB 10; Length 1023;  
Best Local Similarity 72.0%; Pred. No. 7.5e-06;  
Matches 77; Conservative 0; Mismatches 30; Indels 0; Gaps 0;  
QY 1449 TTATTAACAATCATTAACCTGTACACGGTAAGAAATACCTAGGAATAAGTAAGAACT 1508  
Db 896 TTATTTCCAAACATGCTCTCTACTTGAAGTAAAGAAATTAATAGAACTTT 955  
QY 1509 CTTGAGTAAAAA... 1555  
Db 956 TGTGTNA... 1002

## RESULT 10

US-09-925-300-464  
; Sequence 464, Application US/09925300  
; Patent No. US20020151681A1  
; GENERAL INFORMATION:  
; APPLICANT: Craig Rosen,  
; APPLICANT: Steve Ruben,  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA101  
; CURRENT APPLICATION NUMBER: US/09/925,300  
; CURRENT FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05988

; FILE REFERENCE: 2000-03-08  
; CURRENT APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1999-03-12  
; PRIOR FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 1890  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 464  
; LENGTH: 2431  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-300-464

Query Match 3.8%; Score 59.4; DB 10; Length 2431;  
Best Local Similarity 77.4%; Pred. No. 1.4e-05;  
Matches 72; Conservative 0; Mismatches 21; Indels 0; Gaps 0;  
QY 1463 TTAACCTGTACACGTAAGAAATACCTAGGAATAAAGTAAAGCTCCTCAGTAAACAAA 1522  
DB 2337 TTGGTTTGTATCTGTAGCTCAGTACGTCTCTAATAAGCTTAAGATCCTCAAAAAAAA 2396  
QY 1523 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 1555  
DB 2397 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 2429

RESULT 11  
US-10-003-295-1  
; Sequence 1, Application US/10003295  
; Patent No. US20020168741A1  
; GENERAL INFORMATION:  
; APPLICANT: GAN, Weiniu et al.  
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC  
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES  
; FILE REFERENCE: CL001183DIV  
; CURRENT APPLICATION NUMBER: US/10/003,295  
; CURRENT FILING DATE: 2001-12-06  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 2674  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-003-295-1

Query Match 3.8%; Score 59; DB 9; Length 2674;  
Best Local Similarity 55.0%; Pred. No. 1.9e-05;  
Matches 116; Conservative 0; Mismatches 95; Indels 0; Gaps 0;  
QY 1345 CAGGAGTCGCTGGACATGCGGATGAGGTCTTCGGAGTGGAGTGGCGCCCGGATGGCTCT 1404  
DB 2439 CAGCATCCACACTGCGCGCAGGATGCAGCGCGCTGTCTCTGTGTCTCTGTGTCTGTGTCT 2498  
QY 1405 AGAGTTGCTCTGTGTGGCAAGACAAAGTTATAAGCTATGGCTTATTAACAAATCATT 1464  
DB 2499 AGGCTTCTCTTCCGGCAGCAACATAAACCACCTTGTGCCACTCAAAAAAATAAAAA 2558  
QY 1465 AACTGTACACGGTAAGAAATACCTAGGAATAAAGCTCCTCAGTAAACAAA 1524  
DB 2559 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 2618  
QY 1525 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 1555  
DB 2619 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 2649

RESULT 12  
US-09-918-995-30837  
; Sequence 30837, Application US/09918995  
; Publication No. US20030073623A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
; FROM VARIOUS CDNA LIBRARIES

; FILE REFERENCE: 20411-756  
; CURRENT APPLICATION NUMBER: US/09/918,995  
; CURRENT FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US/09/235,076  
; PRIOR FILING DATE: 1999-01-20  
; NUMBER OF SEQ ID NOS: 38054  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 30837  
; LENGTH: 453  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(453)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-918-995-30837

Query Match 3.8%; Score 58.4; DB 9; Length 453;  
Best Local Similarity 48.8%; Pred. No. 1e-05;  
Matches 158; Conservative 0; Mismatches 166; Indels 0; Gaps 0;  
QY 1019 AGGTGGAGTCGCTGGTTTCTGTTCCGGATGACACACCCCTCTATCTGTGCGGAACACC 1078  
DB 60 ACGCGAGTTTGTAGCTCTGTTCCAGGACACAAACATCAAGCTCTGGGACATCAGGA 119  
QY 1079 AGAACAAAGTCGTTGAGCGCATGCACAGGCGCACAGGCGGTGTCACGATGTGAATATT 1138  
DB 120 GGAAGGCTGTCTTCCGATACAGGGGGGACAGCGCGTGTCTCCGGTTCA 179  
QY 1139 GCGCGGATGAAGCTAATTCGCTGCTTTCATTTGACAGTCAGTCCGTCTGTGCGGAG 1198  
DB 180 GCCCGGATGGGAGTGGTGGCGTGGCGGCGAGATGACACACCGTGAAGCTCTGGGATC 239  
QY 1199 CCAGCGATGTCATCATGCGCACCTTCCGGGGTTCATGTCAGGCTGTTTACACGGTTG 1258  
DB 240 TCACTGCGCGCAAGATGATGTCGTAGTTCCCTGTCACAGGGCGCTGTCAACGTTGTCG 299  
QY 1259 CTTGTCGCGGGACTCCCGCTGATTGTTTCCGGCAGCAAGACTCAACTCTAAAAAGTAT 1318  
DB 300 AGTTTACCCACAGGATACCTCTGGCTCCGGCAGCTCTGACGAGCAATCCGCTTCT 359  
QY 1319 GGAGTGTGCAGACGAGAAACATGG 1342  
DB 360 GGGACCTGGAGAGTTCCAGGTGG 383

RESULT 13  
US-09-997-279-19  
; Sequence 19, Application US/09997279  
; Publication No. US20030059781A1  
; GENERAL INFORMATION:  
; APPLICANT: Chenault, Ruth A.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
; FILE REFERENCE: 210121.501C1  
; CURRENT APPLICATION NUMBER: US/09/997,279  
; CURRENT FILING DATE: 2001-11-28  
; NUMBER OF SEQ ID NOS: 230  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 19  
; LENGTH: 1602  
; TYPE: DNA  
; ORGANISM: Homo sapien  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(1602)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-997-279-19

Query Match 3.7%; Score 57.6; DB 9; Length 1602;  
Best Local Similarity 65.1%; Pred. No. 3.3e-05;  
Matches 84; Conservative 0; Mismatches 45; Indels 0; Gaps 0;



OM nucleic - nucleic search,, using sw model  
Run on: June 2, 2003, 02:57:24 ; Search time 2084 Seconds  
(without alignments)  
12084.444 Million cell updates/sec

Title: US-09-830-980-2  
Perfect score: 1555  
Sequence: 1 aattcccaaaaatgcagga.....aaaaaaaaaaaaaaaaaaaa 1555

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues  
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:\*

- 1: em\_estba:\*\*
- 2: em\_esthum:\*\*
- 3: em\_estin:\*\*
- 4: em\_estum:\*\*
- 5: em\_estov:\*\*
- 6: em\_estopl:\*\*
- 7: em\_estro:\*\*
- 8: em\_htc:\*\*
- 9: gb\_est1:\*\*
- 10: gb\_est2:\*\*
- 11: gb\_htc:\*\*
- 12: gb\_est3:\*\*
- 13: gb\_est4:\*\*
- 14: gb\_est5:\*\*
- 15: gb\_estfun:\*\*
- 16: em\_estom:\*\*
- 17: gb\_gss:\*\*
- 18: em\_gss\_hum:\*\*
- 19: em\_gss\_inv:\*\*
- 20: em\_gss\_pln:\*\*
- 21: em\_gss\_vrt:\*\*
- 22: em\_gss\_fun:\*\*
- 23: em\_gss\_man:\*\*
- 24: em\_gss\_mus:\*\*
- 25: em\_gss\_other:\*\*
- 26: em\_gss\_pro:\*\*
- 27: em\_gss\_rod:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	735.6	47.3	790	9	AI518473 LD38001.5
2	597.8	38.4	637	13	BI637470
3	551.8	38.1	725	12	BF505998
4	578.2	37.2	622	13	BI632742
5	549.2	35.3	606	13	BI232464
6	548.8	35.3	626	13	BI639652

7	488.2	31.4	527	9	AI520108
8	485.2	31.2	552	13	BI635777
9	476.6	30.6	1784	11	BC035421
10	455.8	29.3	510	13	BI635228
11	424	27.3	448	9	AA735870
12	346	22.3	688	13	BM579586
13	331.6	21.3	975	9	AL520339
14	314	20.4	925	9	AL526575
15	301.4	19.4	1114	9	AL530778
16	286.4	19.1	861	9	AL558947
17	280.4	18.7	851	9	AL558453
18	289.4	18.6	861	9	AL519702
19	284.2	18.3	997	14	BQ070427
20	282	18.1	927	12	BG470097
21	278.8	17.9	715	13	BM579239
22	278.8	17.9	939	14	BQ956834
23	276.6	17.8	826	9	AL526922
24	275.2	17.7	822	9	AL550330
25	268.8	17.3	1103	13	BM478420
26	265.2	17.1	1006	13	BM474111
27	263.6	17.0	816	9	AL527738
28	263	16.9	573	13	BM654928
29	262.4	16.9	324	13	BI232484
30	256.6	16.5	921	13	BI833055
31	251.4	16.2	912	9	AL522417
32	250.2	16.1	903	12	BG481047
33	245.4	15.8	673	13	BM654934
34	242.6	15.6	678	14	BQ257262
35	242.6	15.6	776	12	BG323761
36	242	15.6	846	13	BI116063
37	241.2	15.5	1080	13	BM467429
38	239.4	15.4	756	12	BG386139
39	239	15.4	897	14	BQ727393
40	238.4	15.3	697	13	BI855975
41	237.4	15.3	843	12	BE792381
42	236.2	15.2	814	13	BI772615
43	233.6	15.0	759	12	BF536680
44	232.6	15.0	274	9	AA696993
45	229.8	14.8	668	14	BM849233

ALIGNMENTS

RESULT 1  
AI518473  
LOCUS AI518473 790 bp mRNA linear EST 23-APR-2001  
DEFINITION LD38001.5prime LD Drosophila melanogaster embryo pDT2 Drosophila melanogaster cDNA clone LD38001.5 similar to Ntc: FBan0002863 'signal transduction' located on: 2L 21C6-21C6; : 04/10/2001, mRNA sequence.  
ACCESSION AI518473  
VERSION AI518473.2 GI:13769352  
KEYWORDS EST.  
SOURCE fruit fly,  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE 1 (bases 1 to 790)  
AUTHORS Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G., Lewis,S. and Rubin,G.M.  
TITLE BDGP/HMI Drosophila EST Project  
JOURNAL Unpublished (2001)  
COMMENT On Mar 16, 1999 this sequence version replaced gi:4424327.  
Other\_ESTs: LD38001.5prime  
Contact: Stapleton, M.  
BDGP  
Lawrence Berkeley National Lab  
One Cyclotron Rd, Berkeley, CA 94720, USA  
Fax: 510 486 6798  
Email: <http://www.fruitfly.org/EST/estefruitfly.berkeley.edu>  
hit genomic AE003589: arm:2L [301580,604052]







```
COMMENT Contact: Stapleton, M.
BDGP Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
hit genomic AE003589: arm:2L [301580,604052]
estimated-cyto:21B7-21C7: 05/23/2001
Plate: SD.266 row: A column: 5
High quality sequence stop: 499.
Location/Qualifiers
1. 622
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="SD Drosophila melanogaster Schneider L2 cell
culture pot2"
/lab_host="DH5-alpha"
/note="Vector: pOT2; Site_1: EcoRI; Site_2: XhoI; Sized
fractionated cDNAs were directly ligated into pOT2.
Plasmid cDNA library."
Plasmid cDNA library."
159 c 188 g 113 t
BASE COUNT 162 a 159 c 188 g 113 t
ORIGIN
Query Match 37.2%; Score 578.2; DB 13; Length 622;
Best Local Similarity 98.9%; Pred. No. 2.6e-84;
Matches 604; Conservative 0; Mismatches 3; Indels 4; Gaps 2;
QY 7 CAAAATGTCAGGAGACGGACACGGAGCAAGAGCCACGATACGATACAGCGCGC 66
DB 13 CAGAAATGTCAGGAGACGGACACGGAGCAAGAGCCACGATACGATACAGCGCGC 72
QY 67 CTCGTTT---ACAGCGGAGGAGCGCGCGCCCAATCGACCTGCCGCGAGGAATCACT 123
DB 73 CTCGTTTGGACACAGCGGAGGAGCGCGCGCCCAATCGACCTGCCGCGAGGAATCACT 132
QY 124 ACCAGCAATTCGGGACTGATTTGCAACGCGCTGCTGAAAACAGAGCAAGCCACTCCATAT 183
DB 133 ACCAGCAATTCGGGACTGATTTGCAACGCGCTGCTGAAAACAGAGCAAGCCACTCCATAT 192
QY 184 TTGTTTTCGTGGCGGAGGATGAGATCAAGAGAGCGCTGGAGACACGTTGCGATTG 243
DB 193 TTGTTTTCGTGGCGGAGGATGAGATCAAGAGAGCGCTGGAGACACGTTGCGATTG 252
QY 244 TCAGTGACACCCGAAACGTCGATATTGTTATTCAGCCAGCGCGGTTTCAAAGTG 303
DB 253 TCAGTGACACCCGAAACGTCGATATTGTTATTCAGCCAGCGCGGTTTCAAAGTG 312
QY 304 CGCCAGTGACAGATGCACGAGTTCCATGCGCGGACACCGCGAGGCTGTGTTCCGCTG 363
DB 313 CGCCAGTGACAGATGCACGAGTTCCATGCGCGGACACCGCGAGGCTGTGTTCCGCTG 372
QY 364 AATTTACGCCGATGGTCTCATCTGCCAGTGGAGTGGCGACACACAGTGGCGATTG 423
DB 373 AATTTACGCCGATGGTCTCATCTGCCAGTGGAGTGGCGACACACAGTGGCGATTG 432
QY 424 TGGATCTTTAACACAGACACCGCCTTCACCTGCGACAGTTCATACAGTGGGTTCTG 483
DB 433 TGGATCTTTAACACAGACACCGCCTTCACCTGCGACAGTTCATACAGTGGGTTCTG 492
QY 484 TGGATCTTCCTGGGCTCCGGATGGCAACGTTGGCCAGCGGTTGCAAGCGGGCTCTATA 543
DB 493 TGGATCTTCCTGGGCTCCGGATGGCAACGTTGGCCAGCGGTTGCAAGCGGGCTCTATA 552
QY 544 ATCATCTGGGACCGGAGACGGGTACGACGAGGGGCGCCCTTGAGTGGGCACAGAAA 603
DB 553 ATCATCTGGGACCGGAGACGGGTACGACGAGGGGCGCCCTTGAGTGGGCACAGAAA 611
QY 604 CACATCAACTG 614
DB 612 CACATCAACTG 622
```

---

```
RESULT 5
B1232464
LOCUS
DEFINITION RE28535:Prime RE Drosophila melanogaster normalized Embryo pf1c-1
Drosophila melanogaster cDNA clone RE28535 5 similar to Nle:
Fnan002863 'signal transduction' located on: 2L 21C6-21C6;:
04/12/2001, mRNA sequence.
ACCESSION B1232464
VERSION B1232464.1 GI:14700036
KEYWORDS EST
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 606)
AUTHORS Stapleton, M., Brokstein, P., Hong, L., Tyler, D., Berman, B., Carlson
, J., Champe, M., Chavez, C., Dorsett, V., Farfan, D., Frise, E., George
, R., Gonzalez, M., Guarin, H., Harris, N., Li, P., Liao, G., Misra, S.,
Mungall, C.J., Nuno, J., Pacleb, J., Paragas, V., Park, S.,
Phonananavong, S., Wan, K., Yu, C., Lewis, S.E., Celnikier, S. and Rubin
, G.M.
TITLE BDGP/HMI RE Drosophila EST Project
JOURNAL Unpublished (2001)
COMMENT Contact: Stapleton, M.
BDGP Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
hit genomic AE003589: arm:2L [301580,604052]
estimated-cyto:21B7-21C7: 04/12/2001
Plate: RE.285 row: C column: 11
High quality sequence stop: 546.
Location/Qualifiers
1. 606
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RE Drosophila melanogaster normalized Embryo
pf1c-1"
/sex="male and female"
/dev_stage="0-24 hours mixed stage embryonic"
/lab_host="DH5-alpha Tona"
/note="Organ: embryo; Vector: pF1c1; Site_1: XhoI; Site_2:
BamHI; Library was kindly generated by Piero Carninci at
the RIKEN The library was normalized and excised using
Cre recombinase. Plasmid cDNA library."
BASE COUNT 150 a 153 c 185 g 118 t
ORIGIN
Query Match 35.3%; Score 549.2; DB 13; Length 606;
Best Local Similarity 98.9%; Pred. No. 1.3e-79;
Matches 564; Conservative 0; Mismatches 3; Indels 3; Gaps 1;
QY 7 CAAAATTCAGGAGACGGACACGGAGCAAGAGCCACGATACGATACAGCGCGC 66
DB 37 CAGAAATTCAGGAGACGGACACGGAGCAAGAGCCACGATACGATACAGCGCGC 96
QY 67 CTCGTTT---ACAGCGGAGGAGCGCGCGCCCAATCGACCTGCCGCGAGGAATCACT 123
DB 97 CTCGTTTTCGACACGGCGAGGAGCGCGCGCCCAATCGACCTGCCGCGAGGAATCACT 156
QY 124 ACCAGCAATTCGGGACTGATTTGCAACGCGCTGCTGAAAACAGAGGAGCCACTCCATAT 183
DB 157 ACCAGCAATTCGGGCTGCTGATTTGCAACGCGCTGCTGAAAACAGAGGAGCCACTCCATAT 216
QY 184 TTGTTTTCGTGGCGGAGGATGAGATCAAGAGAGCGCTGGAGGACAGCTGGGACTTGGCG 243
DB 217 TTGTTTTCGTGGCGGAGGATGAGATCAAGAGAGCGCTGGAGGACAGCTGGGACTTGGCG 276
QY 244 TCAGTGACACCCGAAACGTCGATATTGTTATTCAGCCAGCGGTTTCAAAGTG 303
```

Db	277	TCAGTGGACACGAAACGCTGATCGATATATGTGTATCAGCCACAGCGCGTTCCTCAAGATG	336
QY	304	CGCCCACTGACAAAGATGACACGAGTTCCATGCGGGACACCGCAGGCTGTGGTTTCGCTG	363
Db	337	CGCCCACTGACAAAGATGACACGAGTTCCATGCGGGACACCGCAGGCTGTGGTTTCGCTG	396
QY	364	AAATTTCAGCCCGGATGGTCTCACTCGCCAGTGGAGTGGCGACACACACAGTGGGTTCTG	423
Db	397	AAATTTCAGCCCGGATGGTCTCACTCGCCAGTGGAGTGGCGACACACAGTGGGTTCTG	456
QY	424	TGGGATCTTAAACACAGACACCGCACCTTCACCTGCACAGGTCATGAAGCAGTGGGTTCTG	483
Db	457	TGGGATCTTAAACACAGACACCGCACCTTCACCTGCACAGGTCATGAAGCAGTGGGTTCTG	516
QY	484	TGCGTATCTCTGGCTCCGATGGCAACGTTGGCCAGCGTTGCAAAACGGGCTCTATA	543
Db	517	TGCGTATCTCTGGCTCCGATGGCAACGTTGGCCAGCGTTGCAAAACGGGCTCTATA	576
QY	544	ATCATCTGGGACCCGAGCGGTGACAG	573
Db	577	ATCATCTGGGACCCGAGCGGTGACAG	606
RESULT 6			
BI639652			
LOCUS		626 bp mRNA linear EST 10-SEP-2001	
DEFINITION		SD22291.5prime SD Drosophila melanogaster Schneider L2 cell culture	
		pot2 Drosophila melanogaster cDNA clone SD22291 5 similar to Nile;	
		FBan0002863 GO:[signal transduction (GO:0004871)] located on: 2L	
		21C6-21C6;: 05/19/2001, mRNA sequence.	
ACCESSION		BI639652	
VERSION		BI639652.1 GI:15541862	
KEYWORDS		EST.	
SOURCE		fruit fly.	
ORGANISM		Drosophila melanogaster	
		Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;	
		Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;	
		Ephydroidea; Drosophilidae; Drosophila.	
REFERENCE		1 (bases 1 to 626)	
AUTHORS		Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,	
		Lewis,S. and Rubin,G.M.	
TITLE		BDGP/HMI Drosophila EST Project	
JOURNAL		Unpublished (2001)	
COMMENT		Contact: Stapleton, M.	
		BDGP	
		Lawrence Berkeley National Lab	
		One Cyclotron Rd, Berkeley, CA 94720, USA	
		Fax: 510 486 6798	
		Email: <a href="http://www.fruitfly.org/EST">http://www.fruitfly.org/EST</a> , <a href="mailto:est@fruitfly.berkeley.edu">est@fruitfly.berkeley.edu</a>	
		hit genomic AE003589; arm:2L [301580,604052]	
		estimated-cyto:21B7-21C7; 05/19/2001	
		Plate: SD.222 row: H column: 7	
		High quality sequence stop: 544.	
FEATURES		Location/Qualifiers	
source		1. 626	
		/organism="Drosophila melanogaster"	
		/db_xref="taxon:7227"	
		/clone="SD22291"	
		/clone_lib="SD Drosophila melanogaster Schneider L2 cell	
		culture pot2"	
		/lab_host="DH5-alpha"	
		/note="vector: pot2; Site_1: EcoRI; Site_2: XhoI; Sized	
		fractionated cDNAs were directly ligated into pot2.	
		Plasmid cDNA library."	
BASE COUNT		155 a 158 c 188 g 124 t 1 others	
ORIGIN			
Query Match		35.38; Score 548.8; DB 13; Length 626;	
Best Local Similarity		98.94; Pred. No. 1.5e-79;	
Matches 563; Conservative		0; Mismatches 3; Indels 3; Gaps 1;	
QY	7	CAAAAATGCAGGACGGACCGGACGAGGCGCCACATAGATACAGGCGCGC	66





1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

Oy 32

```

|||||
301 ACAGATTCATCGCGGACACGCGGAGGCTGTGGTTTCGTAATTCACGCCGGATGGT 360
|||||
382 GTCATCTCCGACAGTGGAGTGGGACACACAGTGGCGGATGTGGGATCTTAACACAGAG 441
|||||
361 GTCATCTCCGACAGTGGAGTGGGACACACAGTGGCGGATGTGGGATCTTAACACAGAG 420
|||||
442 ACACCGCACTTCACCTGCACAGGTGCTATA 469
|||||
421 ACACCGCACTTCACCTGCACAGGTGCTATA 448
|||||

RESULT 12
BM579586
LOCUS
DEFINITION
17000687239580 A.Gam.ad.cdna.linear EST 22-FEB-2002
19600449699487 5', mRNA sequence.
BM579586
BM579586.1 GI:18868053
EST.
African malaria mosquito.
ORGANISM
Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
Anopheles.
REFERENCE
1 (bases 1 to 688)
Holt R.A., Lin J.-J., Murphy S.D., Evans C.A., Kraft C.L., Charlab
R., Collins F.H., Venter J.C. and Hoffman S.L.
Celera Anopheles gambiae EST project
Unpublished (2002)
Contact: Holt R.A.
Celera Genomics
45 W. Gude Dr., Rockville, MD 20850, USA
Tel: 2404533151
Fax: 2404534580
Email: HoltRA@celera.com
Plate: NU01004AB3 row: D column: 01
Seq primer: M13 Reverse.
FEATURES
source
Location/Qualifiers
1..688
/organism="Anopheles gambiae"
/strain="RSP-ST (Reduced susc. to Permethrin - std.
chromosome)"
/db_xref="taxon:7165"
/clone="19600449699487"
/clone_lib="A.Gam.ad.cdna.bloodi"
/dev_stage="Adult"
/lab_host="DH10B"
/note="Vector: pSport1; Site_1: SalI; Site_2: NotI; Whole
adult mosquitoes (mixed sex) frozen on liquid nitrogen 24
hours after human blood feeding. cDNA inserts >500 bp
cloned directionally into pSport 1. Not 1 site is 3'.
Clones available through the Malaria Research and
Reference Reagent Resource Center (www.malaria.mr4.org)"
BASE COUNT 136 a 178 c 240 g 134 t
ORIGIN

Query Match 22.3%; Score 346; DB 13; Length 688;
Best Local Similarity 71.0%; Pred. No. 9e-47;
Matches 473; Conservative 0; Mismatches 190; Indels 3; Gaps 1;

QY 793 TCCAAAGATCGCACAGTCAAGATGTGGCGAGACGCTGATGGAATCTTGTGCGCGGACGTTTC 852
|||||
DB 14 TCCGGCGATCGGACGCTCAAGATGTGGCGAGCGGAGACGCGGTACTGTGCAAAACGTTTC 73
|||||
QY 853 TCTGGCCAAAGCTCACTGGGTAAACAACATTCGCTGAGCACCGCATTTAGTCTCTCGGCACCT 912
|||||
DB 74 ACGGGTTCACGCGCACTGGGTGAATAATTTGGCGCTCAACACGCACTAGTCTCTCGGCAGC 133
|||||
QY 913 GTTCCATTCCTCCGCTGAAGATCGCTCCCAAGACGAC--CTCAGTTTCAGACACTGAG 969
|||||
DB 134 GGCCCGTTTACCCGGTGATGACAAATTTAAGATGTACGGGACGCGCGGATGAAGGAC 193
|||||

```

```

QY 970 GAATTGAGGAATCTGCCTTGAAGCGCTACCAGCGCTTGCCTGACGAGCTGGAGTCG 1029
|||||
DB 194 GCCCTCAAGCAGAGTGGCGCTCGAGCGGTACGAGCAGGTGTGCCGCGACGGGTGAATCG 253
|||||
QY 1030 CTGGTTTCCTGTTCCGATGATCAACACCCCTCTATCTGTGGCGGGAACACAGCAAGTGC 1089
|||||
DB 254 TTCGTCTCGTGTTCGGACACCTTTACGCTGTATCTGTGGCGCGCAACACAGAACTTC 313
|||||
QY 1090 GTTGAAGCGCATGACAGGCGCACAGAACGCTGGTCAACAGATGTGAAATATTTCGCCGATGTA 1149
|||||
DB 314 GTGACGCGCATGACCGCGCACCAAGATGTGGTGAACGATGTGAAGTACTCGCGGACCTG 373
|||||
QY 1150 AAGCTAATTTGGCTCTGCTTCATTTGACAAGTCAGTGGCTGTGTGGCGACCGCAGCATGGT 1209
|||||
DB 374 AAGTTTATTTGGCTCGCTTCGTTCCGACAAATCGATCCGGGTGTGGCGCGCGCATGGG 433
|||||
QY 1210 CAGTACATGCGCACCTTCCGGGGTGTGTGAGGCTGTGTACACGGTGTGCTGTGGTCCCGG 1269
|||||
DB 434 CGGTTCATTTGTGCTTGGCGGTGTCACGCTGCTGCGCGTCTACACGGTGAGCTGGTCGGCC 493
|||||
QY 1270 GACTCCCGCTTGATTTTCCGGCAGCAAGACCACTCAACTCTAAAGATATGAGTGTGCAG 1329
|||||
DB 494 GATTCGGCGGTGTAGTGTCTAGCGGACGAGGACCAACCTCAAGGTGTGGAGTGTGAAG 553
|||||
QY 1330 ACGAAGAACTGCGACAGGAGCTGCTGACATGCGGATGAGTGTTCGAGTGGAGCTGG 1389
|||||
DB 554 GAGCGCAAGCTGGTGGCAGGAGTGCCTGCGGACACGCGGAGTGTACGCGCTGGATTGG 613
|||||
QY 1390 CGGCGCGATGCTAGAGTGCCTCTGTGGCAGGACAAAGTATTAAGTATTTAGCTATGGCT 1449
|||||
DB 614 GCACCGGACGGTCCCGGCTTCTCCGGTGGCAAGATAGGTGTAAACTGTGGGCA 673
|||||
QY 1450 TATTAA 1455
|||||
DB 674 TATTAA 679
|||||

RESULT 13
LOCUS
DEFINITION
AL520339 LTI_NFL004_NBC2 Homo sapiens cdna clone CS0DB006YB09 5
prime, mRNA sequence.
ACCESSION
AL520339
VERSION
AL520339.1 GI:12783832
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 975)
Li W.B., Gruber C., Jessee J. and Polayes D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
Location/Qualifiers
1..975
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DB006YB09"
/clone_lib="LTI_NFL004_NBC2"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dN) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed
by Life Technologies. Contact : Feng Liang Life
Technologies, a division of Invitrogen 9800 Medical Center

```







Tue Jun 3 11:27:50 2003

us-09-830-980-2\_1.rst

Page 12

Search completed: June 2, 2003, 04:48:17  
Job time : 2090 secs

---